INVESTIGATION OF GENOMIC ESTIMATED BREEDING VALUES AND ASSOCIATION METHODOLOGIES USING BAYESIAN INFERENCE IN A NELLORE-ANGUS CROSSBRED POPULATION FOR TWO TRAITS

A Dissertation

by

LAUREN LORENE HULSMAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	David G. Riley
Committee Members,	Dorian J. Garrick
	Clare A. Gill
	James O. Sanders
Head of Department,	H. Russell Cross

August 2013

Major Subject: Animal Breeding

ABSTRACT

The objectives of this study were to 1) evaluate marker associations for genomic regions of interest and significant ontology terms, 2) evaluate and compare 4 models for their efficacy in predicting genetic merit, 3) evaluate and compare the impact of using breed-of-origin genotypes in a Bayesian prediction model, and 4) evaluate the effects of data partitioning using family structure on predictions. Nellore-Angus F_2, F_3 and half-sibling calves were used with records for overall temperament at weaning (OTW; a subjective scoring system; n = 769) and Warner-Bratzler shear force (WBSF; a measure of tenderness; n = 389). After filtering, 34,913 markers were available for use. Bayesian methods employed were BayesB (using $\hat{\pi}$) and BayesC (using $\pi = 0$ and $\hat{\pi}$) in GenSel software, where, after estimation, $\hat{\pi}$ = 0.995 or 0.997 for WBSF or OTW, respectively. No regions associated with either trait were found using $\hat{\pi}$, but when $\pi = 0$ associated regions were identified (37 and 147 regions for OTW and WBSF, respectively). Comparison of genomic estimated breeding values from these 3 Bayesian models to an animal model showed that BayesC procedures (using $\hat{\pi}$) had the highest accuracy for both traits, but that BayesB had the lowest indication of bias in either case. Using a subset of the population (n = 440), genotypes based on the breed in which the alleles originated from (i.e., breed-of-origin genotypes) were assigned to markers mapped to autosomes (n = 34,449), and incorporated into prediction analyses using BayesB ($\hat{\pi}$ = 0.997) with or without nucleotide-based genotypes. In either case, there was an increase in accuracy when breed-of-origin genotypes were incorporated into prediction analyses.

Data partitions based on family structure resulted in 13 distinct training and validations groups. Relationship of individuals in the training with validation individuals did have an impact in some cases, but not all. There was poor prediction of genomic estimated breeding values for individuals in the validation population using BayesB methods, but performed better in all cases than breeding values generated using an animal model. Future studies incorporating breed-of-origin genotypes are of interest to determine if accuracy is improved in these groups.

DEDICATION

To my family, close friends, and the love of my life

"Those who hope in the Lord will renew their strength. They will soar on wings like eagles; they will run and not grow weary, they will walk and not be faint."

Isaiah 40:31

ACKNOWLEDGEMENTS

I would be remiss not to acknowledge the many people that have mentored and supported me through my graduate education. To all those that have contributed in some way to my educational pursuits, I sincerely thank you.

My advisor and committee chair, Dr. David Riley, has provided invaluable opportunities, time, and knowledge in my graduate education and in my career pursuits. The experiences and networks I have built due to his generosity and sincere care to enhance my graduate career cannot be repaid. Thank you not only for the experiences, opportunities, and knowledge, but also for setting an excellent example of mentorship.

My committee members, Drs. Dorian Garrick, Clare Gill, and Jim Sanders, provided invaluable knowledge and guidance related to my research as well as life. I'm also greatly appreciative for the opportunity to pursue the teaching aspect of my career, which has generated a strong passion for teaching agriculture within me. To all of you, thank you from the bottom of my heart.

The McGregor Genomics Herd Project has been an integral part to my graduate career at Texas A&M University. To all those involved with this project, thank you for providing an avenue for students like myself to gain experience and knowledge with cattle and in the field of genetics.

Research cannot proceed without funding, so I wish to express great appreciation for the funds that supported my project and graduate pursuits. This project was partially supported by National Research Initiative competitive grant no. 2008-35205-18767 from

the USDA National Institute of Food and Agriculture Animal Genome Program, by Texas AgriLife Research, and the Beef Checkoff.

Last, but definitely not least, I must recognize and thank my support system throughout my graduate career. My family, especially my parents, have financially and emotionally encouraged me to aim for the stars and provided means for me to obtain those goals. My close friends provided reassurance and an ear to listen when needed. The love of my life, although a recent addition, has been a steady rock for me to lean on and provided invaluable encouragement and support. To all of you, thank you! I love all of you so much!

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

The use of breeding values in selection practices, in the form of expected progeny differences (EPDs), have become common in the US beef cattle industry. Traditional methods utilize pedigree information to calculate genetic relationships or similarities and produce estimated breeding values (EBV), which are estimates of the genetic merit or 2 times transmitting abilities. With advancements in molecular technology, particularly the availability of dense SNP marker arrays, the ability to generate genomic estimated breeding values (GEBVs) utilizing this molecular information has become a reality. Such procedures on a genome-wide scale were introduced by Meuwissen et al. (2001), including models using a Bayesian framework.

Research on methods to incorporate these procedures, termed genomic prediction, has primarily focused on simulations and real data using purebred populations, particularly dairy, with a few notable exceptions. The collective outcome of this research has shown that the training population structure and size (i.e., population in which marker associations are formed using phenotypes and genotypes) as well as the marker density are important factors for successful implementation. Due to these constraints and their impact on commercial cattle populations in the U.S., research on the use of this methodology for crossbred populations continues even though implementation within certain breed associations has already started.

The Nellore-Angus crossbred population available for this study provides a unique opportunity to investigate certain attributes of the Bayesian framework and prediction of breeding values using genomic information for overall temperament at weaning and Warner-Bratzler shear force after harvest. Therefore, the objectives of this study are:

1) To evaluate marker associations for both traits found in the training phase for genomic regions associated and ontology terms that may elucidate the genomic functions that those markers are identifying;
2) To evaluate and compare 4 models for both traits for their efficacy in predicting genetic merit, including 3 models using a Bayesian framework;
3) To evaluate and compare the impact of using genotypes based on breed from which the allele originated (i.e., breed-of-origin genotypes) in a subset of the population with a Bayesian model (as determined in Objective 2) utilizing the trait with the largest number of available observations (i.e., overall temperament at weaning); and
4) To evaluate the effects of data partitioning using family structure on the consistency of breeding value prediction using a Bayesian model (as determined in Objective 2) utilizing the trait with the largest number of available observations (i.e., overall temperament at weaning).

CHAPTER II

LITERATURE REVIEW

Advances in Breeding Value Prediction

The use of selection criteria for improvement in animal performance is not a novel concept as it has been around as long as humans have raised animals for production purposes. The methods to generate selection criteria through statistical and quantitative means have evolved over the years to produce breeding values, which are equivalent to 2 times the animals' transmitting abilities for any given trait. Early, notable work on the derivation of conditional means and variances for the multivariate normal distribution (Pearson, 1903) followed by the development of maximum likelihood principles for the analysis of variance and variance estimation work (Fisher 1922, 1925) provided one of the first avenues to produce breeding values. Further work was done on quantifying the inbreeding coefficient and genetic relationships of individuals using path coefficient techniques, which were used to create the relationship covariance matrix (i.e., A matrix; Wright, 1921a, 1921b, 1932, 1934), and those techniques were expanded into computational algorithms and models by Lush (1931, 1933). At that time, the statistical procedures assumed that the data came from balanced designs, in other words each block (if there is one) of the design has the same number of experimental units, each treatment occurs the same number of times in the experiment, and the number of times any two treatments occur together in the same block is the same for all pairs of treatments. In reality however, this is usually not the case, especially with livestock data. Work by

Brandt (1933) and Yates (1934) incorporated the use of unbalanced data into statistical methods (as reviewed by Henderson, 1990).

These early contributions influenced the work of Henderson who developed the mixed model equations (Henderson, 1949) that solved for best linear unbiased estimates (fixed effects; Henderson et al., 1959) and predictions (random animal effects; Henderson, 1963). These equations optimized the use of generalized least squares solutions for fixed effects while also minimizing prediction error variances and incorporating relationship information through pedigree and multiple trait evaluation to generate best linear unbiased predictions (BLUP) of breeding values. Not only did this method minimize the prediction error variances, but more importantly, under the assumption of normality, the predictions maximized the expected genetic merit for those individuals selected out of the possible candidates available (Bulmer, 1980; Gianola et al., 1982; Goffinet, 1983). This was a tremendous contribution to the field and it did not assume the fixed effects (β) to be known like many of its predecessors. This method, however, was not widely accepted or used in sire evaluations until the 1970s. By the 1990s, the use of BLUP and the mixed model equations were the standard in dairy and beef cattle evaluations in the U.S. and other countries (Henderson, 1990).

Much of the delay in using the proposed mixed model equations developed by Henderson was attributed to the computational needs and lack of practical methods of estimating variance and covariance parameters. Hartley and Rao (1967) developed a computational method that estimated sampling variances using maximum likelihood estimation with unbalanced data. This method maximized the logarithm of the likelihood

function (L) using iterative methods to estimate the scale parameter (γ, a ratio of unknown fixed effect and residual variance) and elements in β. Later studies by Henderson (1973) and Dempster et al. (1977) found this method to be efficient, but it was noted by Patterson and Thompson (1971) that the use of maximum likelihood incorporated a bias in estimates when the number of fixed effects within β were unequal relative to the observations used and therefore did not have enough degrees of freedom to estimate the fixed effects without bias (see also Falconer and Mackay, 1996). Because of this issue, Patterson and Thompson (1971) introduced a modified version of maximum likelihood, later termed restricted maximum likelihood (REML) because it did not maximize on all parameters at once and, therefore, avoided the bias due to the loss of information from estimating unknown fixed effects.

In their paper, Patterson and Thompson (1971) partitioned the data into two sets, where they first maximized the logarithmic likelihood of the scale parameter (γ and σ_e^2) estimates along with the fixed effects (L') followed by maximizing the logarithmic likelihood of variance component estimates (L''). In this case L = L' + L'', but allowed for the recovery of parameter estimates. In this case, the method gave the same results as the original method proposed by Nelder (1968). When information on selected individuals and their respective progeny were provided, REML had the advantage of estimating the base population parameters, regardless if culled individuals without progeny were not used in the analysis, an important property for solutions through mixed model equations (Henderson, 1990). From there, advances in iterative and data transformation techniques further enhanced the ability to use REML and the mixed

model equations for breeding value solutions. Selection bias was controlled more adequately when Henderson (1976) and Quaas (1976) introduced rapid methods to compute the inverse of the relationship covariance matrix (A^{-1}), which facilitated the incorporation of the animal relationships into the equations. Following these advances, the use of estimated breeding values (EBV) for sire evaluations developed rapidly and was used widely by breed associations, especially dairy, over the next few decades for selection purposes within and across herds.

The modeling and prediction of breeding values through BLUP methodologies relied on pedigree information to ascertain relationship between individuals, which should capture the majority of those animals' genetic relationship. Pedigree estimates of relationships are based on an animal transmitting an average half of its genetic information to its offspring, and, for siblings or non-parent relationships, this probability is carried through the pedigree paths connecting that particular pair of animals. On the genomic side of this relationship, however, a pair of individuals could inherit the same or completely opposite genetic information compared to a single parent or ancestor, meaning that the actual relationship between those two individuals may be less or more compared to pedigree calculations alone (VanRaden, 2007, Goddard, 2009).

Around the same time as advances in statistical genetic evaluations were occurring, advances in genomic mapping and use of DNA markers were also taking place. Many traits are quantitative in nature, being influenced by many, perhaps thousands, of loci in the genome. Fernando and Grossman (1989) demonstrated a method for incorporating markers linked to quantitative trait loci (QTL) into BLUP

breeding values with the idea that this would lead to faster genetic gain and would reduce the generation interval between breeding individuals. Indeed, Meuwissen and Goddard (1996) conducted a simulation study of this technique and predicted that extra genetic gain could be as high as 38% with the inclusion of marker-linked QTL information. At the time, there were limitations preventing this methodology from being incorporated readily into BLUP breeding values because the number of genetic markers available for use was limited, especially those known to be linked with QTL, as were appropriate ways to model linkage for each marker with QTL across families in complex pedigrees. Furthermore, the use of only a few markers for selection on a specific trait relied heavily on the markers accounting for a moderate to large proportion of the genetic variance. In reality, however, the number of markers that account for a large proportion of the genetic variance depends highly on the trait and the density of markers available on the genomic map.

In 2001, Meuwissen, Hayes, and Goddard published a landmark paper that would revolutionize the idea of marker-assisted selection by expanding the use of markers across the entire genome. They proposed the use of dense marker maps to exploit linkage disequilibrium (LD) between the markers and QTL, specifically using haplotypes of markers (i.e. chromosomal segments), and using those associations to predict breeding values. This would capture, conditional on the density of markers, the actual genetic relationship between individuals regardless of relationship based on pedigree. Estimating allelic effects by traditional least squares methods would be troublesome, however, as there would not be enough degrees of freedom to estimate the

effects of thousands of haplotypes or markers in a data set of limited size (Lande and Thompson, 1990; Meuwissen et al., 2001). In their paper, they used a simulation study to compare 4 approaches to dealing with this issue using 1) a stepwise least squares approach that accumulates haplotypes into the model one at a time in order to address the degrees of freedom problem, 2) a BLUP model that fitted all haplotype effects as random and assumed known equal variances per locus, 3) a Bayesian model with all haplotype effects fitted as random and locus-specific variances (BayesA, Table 1), and 4) a Bayesian model with random haplotype effects and locus-specific variances, but modeling a proportion (π) that was assumed to have no association with the QTL and therefore a null (zero) effect (BayesB, Table 1).

Meuwissen et al. (2001) asserted 4 main points from this simulation study: 1) the use of dense markers distributed on all chromosomes could provide accurate estimates of breeding values even if phenotype records are not available or progeny information is limited, 2) the least squares approach could not estimate all haplotype effects simultaneously, and overestimated the effects when included stepwise in the model resulting in lower breeding value accuracy, 3) the Bayesian models that accommodated a prior distribution for the haplotype variances resulted in more accurate predictions (i.e., higher correlations between true, usually simulated values, and predicted breeding values) than the other models even if the prior provided was not correct, and 4) the use of breeding values predicted from markers (in their case haplotypes) for selection, which would come to be termed genomic selection, was considered likely to increase the rate of genetic gain markedly, particularly if techniques were used to shorten the generation

Table 1. Bayesian models developed for genomic selection using Student t-distributions and MCMC iterations

	Model feature						
Model	Probability for a locus to be a QTL (1-π)[1]	Locus-specific variance	Prior distributions[1]	Hyperparameters (estimated)	Hyperparameters (assumed known)	Sampler used	Reference
Bayes BLUP	1	No	$\alpha_i \vert \sigma_a^2 \sim N(0, \sigma_a^2)$, $\sigma_a^2 \sim \chi^{-2}(v, S)$		ν, S	Gibbs	Verbyla et al. (2009)
BayesA	1	Yes	$\alpha_i \vert \sigma_{a_i}^2 \sim N(0, \sigma_{a_i}^2)$, $\sigma_{a_i}^2 \sim \chi^{-2}(v, S)$		ν, S	Gibbs	Meuwissen et al. (2001)
BayesB	1-π	Yes	$\alpha_i \vert \pi, \sigma_{a_i}^2 \begin{cases} 0 \; with \; p(\pi) \\ \sim N(0, \sigma_{a_i}^2) \; with \; p(1-\pi) \end{cases}$, $\sigma_{a_i}^2 \sim \chi^{-2}(v, S)$		ν, S, π	Gibbs/ Metropolis-Hasting	Meuwissen et al. (2001); Solberg et al. (2006, 2008)
Bayes SSVS	1-π	Yes	$\alpha_i \vert \gamma_i, \sigma_{a_i}^2 \sim (1-\gamma_i) N\left(0, \frac{\sigma_{a_i}^2}{100}\right) + \gamma_i N(0, \sigma_{a_i}^2)$, $\sigma_{a_i}^2 \sim \chi^{-2}(v, S)$ $\gamma_i \sim Bernoulli(\pi)$ $1 - p(\gamma_i = 1) = p(\gamma_i = 0) = \pi$		ν, S, π	Gibbs	Yi et al. (2003); Verbyla et al. (2009); Verbyla et al. (2010)
BayesC	1-π	No	$\alpha_i \vert \pi, \sigma_a^2 \begin{cases} 0 \; with \; p(\pi) \\ \sim N(0, \sigma_a^2) \; with \; p(1-\pi) \end{cases}$, $\sigma_a^2 \sim \chi^{-2}(v, S)$ $\gamma_i \sim Bernoulli(\pi)$		ν, S, π	Gibbs	Habier et al. (2010a); Kizilkaya et al. (2010)
BayesCπ	1-π	No	$\alpha_i \vert \pi, \sigma_a^2 \begin{cases} 0 \; with \; p(\pi) \\ \sim N(0, \sigma_a^2) \; with \; p(1-\pi) \end{cases}$, $\sigma_a^2 \sim \chi^{-2}(v, S)$ $\gamma_i \sim Bernoulli(\pi), where$ $\pi \sim Uniform(0,1)$	π	ν, S	Gibbs	Habier et al. (2010a, 2011); Sun et al. (2011)
BayesD	1-π	Yes	$\alpha_i \vert \pi, \sigma_{a_i}^2 \begin{cases} 0 \; with \; p(\pi) \\ \sim N(0, \sigma_{a_i}^2) \; with \; p(1-\pi) \end{cases}$, $\sigma_{a_i}^2 \sim \chi^{-2}(v, S), where$ $S \sim \text{Gamma}(1,1)$	S	ν, π	Gibbs/ Metropolis-Hasting	Habier et al. (2010a, 2011)
BayesDπ	1-π	Yes	$\alpha_i \vert \pi, \sigma_{a_i}^2 \begin{cases} 0 \; with \; p(\pi) \\ \sim N(0, \sigma_{a_i}^2) \; with \; p(1-\pi) \end{cases}$, $\sigma_{a_i}^2 \sim \chi^{-2}(v, S), where$ $S \sim \text{Gamma}(1,1) \; and$ $\pi \sim Uniform(0,1)$	S, π	ν	Gibbs/ Metropolis-Hasting	Habier et al. (2010a, 2011)

Table 1. Continued

Model	Model feature						
	Probability for a locus to be a QTL (1-π)[1]	Locus-specific variance	Prior distributions[1]	Hyperparameters (estimated)	Hyperparameters (assumed known)	Sampler used	Reference
Ante-BayesA	1	Yes	$\alpha_i\{^{\delta_1 \; if \; i=1}_{t_{i,i-1}g_{i-1}+\delta_i \; if \; 2 \leq j \leq m}$ $\delta_i \sim NID\left(0, \sigma^2_{\delta_i}\right), j = 1, \ldots, m$ $\sigma^2_{\delta_i} \sim \chi^{-2}(v, S)$ $t_{i,i-1} \sim N(\mu_t, \sigma^2_t)$		ν, S, μ_t, σ^2_t	Gibbs	Yang and Tempelman (2010, 2012)
Ante-BayesB	1-π	Yes	$\alpha_i\{^{\delta_1 \; if \; i=1}_{t_{i,i-1}g_{i-1}+\delta_i \; if \; 2 \leq i \leq m}$ $\delta_i \sim NID\left(0, \sigma^2_{\delta_i}\right), i = 1, \ldots, m$ $\sigma^2_{\delta_i} = 0 \; with \; p(\pi)$ $\sigma^2_{\delta_i} \sim \chi^{-2}(v, S) \; with \; p(1-\pi)$ $t_{i,i-1} \sim N(\mu_t, \sigma^2_t)$		ν, S, π, μ_t, σ^2_t	Metropolis-Hasting	Yang and Tempelman (2012)

[1] π is the proportion of locus not contributing to the trait of interest; α = estimated substitution effect for the ith marker; σ^2 = marker variance of all markers (constant) or locus-specific (ith marker); ν = degrees of freedom; S = scale parameter, where the two parameters ν and S are used as *a priori* distribution for QTL effect; γ_i = indicator variable (0 or 1) of the ith marker on whether the marker is included in the model based on the probability (p) of π; δ is the ith marker specific effect; and $t_{i,i-1}$ = the marker interval-specific antedependence parameter

interval between breeding individuals. These conclusions were supported through another simulation study by Kolbehdari et al. (2007) using the dairy cattle daughter yield deviation.

Bayesian Methods of Modeling

The concept of using genome-wide markers to find QTL associations with Bayesian methods was introduced in the mid-1990s (Satagopan and Yandell, 1996; Satagopan et al., 1996; Sillanpää and Arjas, 1998; Stephens and Fisch, 1998), but the Bayesian framework provided by Meuwissen et al. (2001) was the first to extend the QTL mapping theory on a genome-wide scale to prediction of breeding values (i.e., genetic merit). Bayesian approaches are flexible in accounting for uncertainties in the data and allow for inferences to be made by averaging across all possible models rather than selecting a single model, meaning more robust inferences than non-Bayesian methods (Meuwissen et al., 2001; Gianola et al., 2003; Xu, 2003a; Yi and Shriner, 2008). This is especially useful in the universal case where the number of markers or segments is greater than the number of observations (Gianola et al., 2003; Xu, 2003a; Yi et al., 2003; Yi and Shriner, 2008). Xu (2003a) summarized the Bayesian framework well, and explained that variables can be grouped into 2 classes: known or observable (e.g. data observed with trait and genotypes of markers) and unknown or unobservable (i.e., parameters). Estimates of unknown variables are from a conditional distribution of the parameters (e.g. means, variances, etc.) based on the priors and the observed data, which is termed the posterior distribution.

This posterior distribution is characterized as the product of the likelihood function, which is a distribution of the observed data based on the unobserved parameters, and the prior distribution, which is the distribution of the unobserved parameters based on *a priori* knowledge. Samples (i.e., random draws from the distribution incorporating the likelihood and prior) are drawn from the entire parameter space (i.e., all possible parameter values) using the joint distribution of the likelihood and prior (i.e., the joint posterior distribution). Although the true posterior distribution is the joint distribution divided by the integration of all possible marginal distributions for the data (i.e., the total distribution), for continuous variables the total distribution often becomes equal to one and therefore the posterior parameter values can be sampled from the joint distribution alone, providing a less cumbersome method.

To produce the posterior means and variances that are of interest, these must be sampled starting with our prior knowledge of the parameters (i.e., the prior distributions), updated after considering the data, and repeated until the best estimate (or the overall mean of the distribution of the samples) for that parameter is obtained using iterative techniques. The entire parameter group (e.g. means and variances together) can be jointly estimated through this iterative process, which differs from some classical approaches that estimate them separately. Because of these properties, the prior distribution and the iterative process can drive the results of the posterior mean and variance estimates (Yi and Shriner, 2008).

With this in mind, Meuwissen et al. (2001) adopted 2 forms of Bayesian methods based on the prior distributions, which were described briefly earlier and are outlined in

Table 1. For both BayesA and BayesB methods, the prior distributions of marker variances are sampled from a scaled inverted chi-square distribution, $\chi^{-2}(\nu, S)$, that is characterized by a scaling parameter (S) that determines the shape of the distribution and the degrees of freedom (ν) that determines the variability of the distribution. In the case of BayesA, all markers are modeled as contributing some genetic variance associated with the trait and therefore each marker has a unique variance estimated. In contrast, BayesB assumes that a proportion of markers (π) have zero effect and therefore adopts a split prior distribution based on the probability of π (see Table 1). To sample the posterior means of the parameters, Meuwissen et al. (2001) employed a Gibbs sampler for both BayesA and BayesB because it utilizes conditional distributions, making it an appropriate choice to sample multiple parameters from various assigned distributions. Due to the split prior distribution employed by BayesB on the marker variance, an additional step incorporating the Metropolis-Hastings acceptance probability was used in the methods proposed by Meuwissen et al. (2001) as the Gibbs sampler could not move throughout the entire sampling space. In both cases, Markov chain Monte Carlo (MCMC) techniques were used to iterate the Gibbs and Metropolis-Hasting steps as it uses the previous chain's estimates within the distributions to sample the current estimates and then moves on without influence from other earlier chains (i.e., the chain is memoryless).

Although there are advantages to the methods described by Meuwissen et al. (2001), much discussion and research has been conducted to investigate the appropriate form of the prior distribution and the sampling techniques used, especially using

regression versus other approaches like semiparametric modeling. This study focuses on Bayesian regression functions, excluding Bayesian LASSO (least absolute shrinkage and selection operator; see Yi and Xu (2008) for description). A review on semiparametric and non-parametric modeling can be found in Gianola et al. (2010).

Following Meuwissen et al. (2001), many of the models that were developed stemmed primarily from the assumptions made about the prior distributions (Table 1), especially with the assumption that certain hyperparameters were known (i.e., degrees of freedom [ν] and scaling parameters [S]) because these influenced the robustness of the solutions. In addition, the ability of the parameters to be drawn from throughout the inference space (i.e., mixing within the sampling space) was limited with the sampling techniques proposed, which could affect the learning process of the sampling algorithm, a key feature in utilizing Bayes theorem (Gianola et al., 2009). Because of this, other iterative techniques utilizing expectation-maximization (EM) algorithms, Iterative Conditional Expectation (ICE) algorithm, and some empirical Bayesian techniques were explored in order to increase computational speed without reducing accuracy of estimates.

Xu (2003a) investigated Bayesian methods to estimate marker effects for barley to conduct QTL association studies for 7 traits. In this study, Xu (2003a) considered using uninformative priors (those distributions that utilize general or vague information about variables to create the sampling distributions) for the BayesA method proposed by Meuwissen et al. (2001), but extended the method across all marker loci instead of haplotypes. Within this study, Xu (2003a) noted that the Bayesian regression method

handled the multicollinearity problem (i.e., association of the same genomic region by multiple markers, which therefore biases the marker effect estimates) and allowed for QTL associations to be identified precisely. The estimates were sampled over many iterations using the Gibbs sampler, which often converges very slowly when parameters are correlated, therefore ter Braak et al. (2005) warned that the uninformative priors that Xu (2003a) adopted could yield improper (i.e., invalid because there is an infinite range that the samples can be pulled from) posterior distributions for the parameter estimates. Although the QTL in the work of Xu (2003a) were identified precisely, ter Braak et al. (2005) was concerned this was an artifact of the convergence issue with the Gibbs sampler and therefore suggested an alternative prior distribution.

Using the same data, ter Braak et al. (2005) extended the individual variance priors used in Xu (2003a) to include a proportion value (δ, where $0 < \delta \leq ½$), where in this case δ is used to keep the prior from having infinite range (i.e., yields a proper prior distribution). Furthermore, ter Braak et al. (2005) eliminated the prior distributions for $\delta = 0$ because it results in an infinite mass of posterior estimates near zero. The new prior variances were sampled from a scaled inverted χ^2 distribution that differed from Meuwissen et al. (2001) because it was characterized by using the proportion value in the exponential form. This improved mixing of variance sampling in some, but not all cases, and is suggested to aid in modeling QTL selection models that have epistatic and genotype-by-environment effects.

Verbyla et al. (2009) introduced a Bayesian version of stochastic search variable selection (SSVS) for prediction of genomic breeding values, which was initially shown

to work well for QTL mapping by Yi et al. (2003). This method allows for a constant dimensionality across all models because it keeps all effects (predictive or not), and reduces non-predictive effects to values very close to zero while also allowing the set of predictive markers to change through iterations. In comparison, the BayesB method utilizes a reverse jump algorithm to drop non-predictive terms before proceeding, therefore making it more computationally demanding on time than the simple Gibbs sampler needed for the Bayes SSVS (Table 1). In addition, the Bayes SSVS introduces a latent indicator variable (γ) for markers or haplotypes that could be either 0 (not included) or 1 (included) and allows relevant information to be extracted so that predictive subset of markers can be found without changing dimensionality, which is different than both BayesA and BayesB. Comparison of correlations between true and predicted genomic breeding values for various dairy traits showed that Bayes SSVS had similar accuracies to the BayesB, had slightly higher accuracies (0.001 to 0.031) than BayesA, and accuracies that ranged from -0.019 to 0.094 greater in comparison to Bayes BLUP (Table 1), which is the same as BayesA in this case, except it assumes that all markers share a constant genetic variance estimated using a Gibbs sampler.

Suggestions by Gianola et al. (2009) spurred development of several other Bayesian models, including BayesC and BayesCπ methods that utilize a proportion of markers (π) not indicative of a QTL but introduces an indicator variable (γ), making it similar to Bayes BLUP except not all markers are included in the model (Habier et al., 2010a, 2011; Kizilkaya et al., 2010). In the case that π is unknown, which is most cases, then BayesCπ could be implemented to identify the appropriate value (Table 1; Habier et

al., 2010a, 2011; Sun et al., 2011). In addition, Habier et al. (2010a, 2011) approached the issue of unknown scale parameters (S) by creating the BayesD and BayesDπ models, where the scale parameter (S) was assumed unknown and sampled from a Gamma distribution. Verbyla et al. (2010) compared 4 models, including Bayes BLUP, BayesA, BayesC (which, by their methods is a Bayes SSVS), and a hybrid of BayesA/BayesB, which was created to decrease computational requirements for their prior distributions. They concluded that all models produced highly correlated genomic estimated breeding values (GEBV), indicating that the hierarchical model that Bayesian analysis adopts by logically moving from one model parameter to the next may be insensitive to the choice of prior distributions for the parameters. Habier et al. (2010a, 2011) also compared their newly developed models of BayesC, Cπ, D, and Dπ with BayesA and B. No large differences in accuracies of GEBV were noted and they concluded that the best modeling method would need to be determined on a trait by trait basis for the data set being investigated.

Meuwissen et al. (2009), Hayashi and Iwata (2010), and Shepherd et al. (2010) expanded on BayesB by trying to decrease computational time through changing the distribution type for marker effects and variances from a Student's t-distribution to using a double exponential (DE) distribution, often called Laplace distribution. The DE distribution has a greater density at zero than a Student's t-distribution which would indicate that more shrinkage can occur, but still resembles a heavy tail distribution similar to a Student's t-distribution. In the original BayesB proposed by Meuwissen et al. (2001), the MCMC iterations using Gibbs and Metropolis-Hasting sampling steps were

computationally demanding while moving between dimensions (i.e., keeping predictive versus dropping non-predictive marker effects). To decrease computational time, the iterative technique was changed in addition to using the DE distribution. Meuwissen et al. (2009) changed the iterative technique to ICE, which uses the expectation or mean instead of the mode of the posterior, which is used in EM algorithms and called the maximum *a posteriori* (MAP) estimate. The computational time was reduced considerably, where the proposed fast BayesB (fBayesB) took 2 to 5 minutes rather than 47 hours for the original, MCMC BayesB. Although accuracies were similar (~0.011 difference) for the two versions of BayesB, it was noted that the fBayesB had a strong conservative behavior so that variances of markers were underestimated, which in turn led to larger bias in the marker effects than MCMC BayesB.

Shepherd et al. (2010) and Hayashi and Iwata (2010) both proposed the use of expectation-maximization algorithms instead of MCMC techniques. In this sense the estimated mode or expected value of the marker variances is found and then the marginal posterior is maximized for all other parameters, usually the marker effects, the residual variance, and any fixed effects. This allows for the MAP estimate or mode to be found without necessarily having the full posterior, meaning a faster computational algorithm (Yi and Banerjee, 2009). Shepherd et al. (2010) expanded on the BayesB model by changing the distribution to DE like Meuwissen et al. (2009), but used the EM iterative technique instead of ICE and termed this new model emBayesB. In their simulations, they found that as the heritability increased for ICE and EM with constant hyperparameter values, the correlation of true to estimated genomic breeding values

decreased. If, however, the emBayesB method also updated the parameters of π, λ (a regularization parameter for the distribution), and the residual variance, then the correlation remained constant (0.87 to 0.88) regardless of heritability. This was also true when generations of animals were split and estimated separately at a heritability of 0.5, and the regression coefficient was close to the original value with the lower heritabilities being overestimated slightly (0.05 to 0.13 difference), which was the lowest across all models compared and indicates less bias in the estimates.

Hayashi and Iwata (2010) used the EM algorithm with a DE distribution on BayesA, but allowed the indicator variable (γ) to be incorporated into the model similar to Bayes SSVS. This model was not a true Bayes SSVS, however, because the mixture distribution of the prior in SSVS cannot be handled well by the EM algorithm. Instead, they weighted each SNP based on its strength of association with the trait using the indicator variable so that a modified version of BayesA using the EM algorithm was used and compared to the original BayesA and BayesB. The weighted BayesA (termed wBSR) performed well (i.e., predicted breeding values had high accuracies) across varying levels of π using simulated data and in some cases performed better than the original BayesA. However, BayesB was more constant in accuracy levels across all π values. Similar to what has been shown before, improvement of the parameters estimates for ν and S improved breeding value accuracies (0.147 increase).

Yang and Tempelman (2010, 2012) proposed an antedependence model using Bayesian methods that exploit the non-stationary correlations between markers and the unknown QTL through LD without the need for haplotypes. This concept was brought

about by the idea that correlations between repeated measures were not always the same depending on the location or time (i.e. non-stationary). Termed Ante-BayesA and Ante-BayesB (Table 1), these models extend the Meuwissen et al. (2001) BayesA and BayesB models to incorporate a non-stationary first-order antedependence correlation structure, which allows the model to change based on location in a specific order (i.e., mapped SNP positions) and tests for significant correlations between successive locations (Yang and Tempelman, 2012). These models still assume locus-specific variances drawn from the same distribution but exploit the LD between markers instead of the haplotypes Meuwissen et al. (2001) used and are proposed to have greater gains in accuracy by doing so. These methods are new to genomic selection, however, and warrant more research into SSVS methods and computational algorithms.

Following the introduction of these Bayesian shrinkage models, concern was raised by Kärkkäinen and Sillanpää (2012) about each model being treated individually rather than creating a framework that illustrates their similarities as well as the method in displaying the components of the model (i.e., priors, likelihoods, and estimators). They described a general framework that formed 4 models that could be altered based on the prior distribution and components; so that the previously described models could be compared on their hierarchical structure (if one exists). In their discussion, they note that the Bayesian models described are powerful tools to estimate effects for genetic markers in both the QTL mapping case and for genomic selection. Furthermore, they have a clear advantage over the animal model in the case that traits are controlled by a limited number of genes, but as the number of genes influencing the trait becomes polygenic,

the animal model is a competitive alternative. Kärkkäinen and Sillanpää (2012) also noted that both the Student's t-distribution and Laplace (i.e., DE) distribution are suitable prior distributions; however, when the trait is polygenic in nature, then the Laplace distribution is more efficient. There were pros and cons for both MCMC and EM sampling techniques, but with the use of DE distributions, it was noted that the EM algorithm may be more easily tuned. Lastly, the use of nonhierarchical models (i.e., fBayesB, emBayesB, and one of their models (Model III), which used a non-hierarchical DE distribution) altered the properties and behavior of the model, which influences the mixing and convergence properties of the estimation algorithm. Kärkkäinen and Sillanpää (2012) noted that the use of hierarchical models reduced the number of variables needed, meaning a more straightforward implementation, faster estimation, and easier and more accurate ability to tune a hyperparameter for the prior distribution (i.e., λ, γ, π, etc).

Model Considerations

Comparisons of ante-Bayes models, Bayes models, and semi-parametric models together are limited and inconsistent across studies. Much more research is needed to understand the benefits of each within specific trait classes and measurement forms. Within each model, however, there are specific considerations to take into account.

Quantitative Trait Loci Distribution. An important consideration is the resemblance of the distribution of SNP or haplotype effects to the underlying QTL distribution (Calus, 2010). Many traits are considered to have distributions where many

loci have small or zero effect on the trait and a limited number have large effects (Hayes and Goddard, 2001), but most are truly unknown and therefore are given a presumed distribution *a priori*. This may lead to biased estimates of QTL effects (Xu, 2003b). The use of single marker associations (usually SNP) rather than haplotypes can often lead to multiple markers explaining the same QTL, however the effect of these markers are usually smaller than the actual QTL effect. Because of this, models should have the capability to make inferences from the data taking into account the underlying distribution of QTL effects. In the Bayesian context, this is accomplished by the prior distribution and the proportion of markers thought (or known) to contribute to the trait of interest (i.e., 1 - π), as these models have greater ability to assign marker effects and variances based on the most predictive set of markers and shrink the rest towards zero.

Estimation of Effects. Marker density across the genome and the marker type are especially important. The original description of Meuwissen et al. (2001) utilized microsatellite haplotypes with the intention of reducing the number of effects compared to the observations available. These haplotypes were important because they yielded many "alleles" per segment, giving greater information over each chromosomal segment with fewer total segments. If these haplotypes were treated as alike-by-state, which accounts for recombination, or identical-by-descent (IBD), which accounts for relationship between haplotypes, the information from using haplotypes was increased even more.

Calus et al. (2009) investigated the haplotype size on the accuracy of QTL mapping compared to prediction of genomic breeding values in a simulation study. The

purpose behind these 2 concepts is often different, as mapping of QTL aims to maximize the contrast in explained variance using the marker interval whereas prediction of breeding values using genomic information aims to maximize the total amount of QTL variance accounted for with the marker intervals. Calus et al. (2009) found that, for prediction of genomic breeding values on young offspring, as the haplotype size increased (e.g., 2, 6, 12, or 20 markers per haplotype) along with stronger IBD probabilities (i.e., more haplotypes available), slightly higher accuracies resulted. However, Calus et al. (2009) found that haplotype variance captured for QTL purposes was often larger with intermediate window sizes (6 to 12 markers per haplotype), indicating that the window sizes and formation of haplotypes should be chosen to increase accuracy based on the goal of the study.

Haplotypes can be considered the same when IBD probabilities are close to 1, which reduces the number of effects that needs to be estimated, increases the power of estimating those effects, and speeds up time to convergence (Yu et al., 2005; Calus et al., 2009). Solberg et al. (2006) investigated the differences between haplotypes and single markers at different densities, where they noted that 4 to 5 times greater density of SNP markers would be necessary to reach comparable prediction accuracy to microsatellites. With SNP densities higher than microsatellite density, however, the use of single marker associations could actually yield higher prediction accuracies (Calus et al., 2007), meaning less computational requirements may be needed to assess all the allelic effects that would typically occur at a single microsatellite locus (or haplotype) and would be more cost effective by today's standards (Garrick, 2010). The issue with single marker

associations rather than haplotypes becomes the strength of LD carried through generations (Villumsen et al., 2009), which Yang and Tempelman (2012) addressed with their proposed Ante-Bayes models.

Piyasatian et al. (2006) investigated using low density marker maps to develop a composite line by exploiting the higher degree of LD created when crossing 2 simulated inbred lines. They found that lower densities of markers performed well based on a calculated cumulative response over their BLUP methodology, but cautioned that the performance was due to the ability to identify marker with the line it originated from, which may not be the case completely with crosses in livestock animals. Habier et al. (2009) also investigated the use of low-density marker panels in an attempt to reduce genotyping costs. Within this study they noted that accuracy of using evenly spaced SNP markers was not affected by the number of QTL with Bayesian approaches, but increases were seen when more of the markers were predictive of QTL making the panel trait dependent. On the other hand, using evenly spaced low density panels derived from high density genotyped ancestors allowed for selection candidates to have a high density panel imputed from their ancestors and was trait independent. Imputation would allow animals to have genotypes inferred for markers not previously genotyped by using ancestor genotypes and statistical probabilities. Low density panels would be more cost effective for producers (Garrick, 2010), and minimum loss of accuracy was seen using this method compared to not deriving high density genotypes for selected individuals, but, in all cases, use of high density panels had higher accuracy and, therefore, outperformed the use of lower density panels (Habier et al., 2009). Moser et al. (2010)

applied this to a Holstein population and concluded that accurate predictions could be achieved with SNP panels of about 3,000 to 5,000 evenly spaced markers. Likewise, Rolf et al. (2010) stated that a minimum of 2,500 and up to 10,000 SNP markers could generate accurate GEBV for commercially important Angus sires. All indicate that smaller marker panels could be used for commercial use in order to reduce costs, but breed-to-breed variation of marker allele frequencies could cause issues with a single SNP commercial panel for evaluation purposes.

Finally, Meuwissen and Goddard (2010) investigated the use of expanding genomic selection methodologies on whole-genome sequencing data and concluded that there was an improvement in prediction accuracy. This would, in essence, utilize an individual's entire DNA sequence to create marker-associations and, therefore, predict breeding values using all information the DNA provided. It is proposed that the cost of sequencing an individual's genome could be less than $1,000 in the near future, and, although this may not be cost effective on a large scale, selected individuals could be sequenced and the remaining population could be genotyped on a high density chip, where their entire genome sequence could be imputed from the selected (most likely ancestral) individuals. Faster, more improved algorithms would need to be created to handle the mass of data this would entail.

Polygenic Effects. The use of marker information, with these methods, is capturing some, if not all, of the relationship between related individuals in a population, but there may be some of the additive genetic variance (i.e., σ_a^2) not captured by the markers. Habier et al. (2007) investigated the relationship captured by markers using 3

different statistical methodologies of fixed regression least squares, random regression-BLUP (RR-BLUP), and BayesB. Markers in linkage equilibrium (LE) were used to determine the genetic relationship captured by markers, where RR-BLUP used all the available LE markers, thereby capturing the most genetic relationship as compared to the other 2 methods. It was noted, however, that the BayesB method had increased accuracy of GEBVs due to different models utilized in the MCMC chains.

Meuwissen et al. (2001) made the assumption that all genetic variation, including relationship between related individuals, was captured by the haplotypes included in the model, therefore, an animal effect (i.e., polygenic effect) was not included in the model. Calus and Veerkamp (2007) argued that this may be a faulty assumption, as not all of the trait variation may be captured by the haplotypes. They investigated the accuracies of GEBV when including and excluding the polygenic effect in the model and found that at high density of markers there was no increase in accuracy and at low density there was marginal increases of accuracy (~1%), when the polygenic effect was included in the model.

Calus and Veerkamp (2007) further concluded that, although no additional accuracy was gained by including the polygenic effect at high density SNP, the estimates of variance components became closer to the original, simulated values and therefore reduced the bias associated by not including the polygenic effect. Rius-Vilarrasa et al. (2012) and Solberg et al. (2009) both supported these conclusions and Solberg et al. (2009) further concluded that inclusion of the polygenic effect over several generations

resulted in slightly higher accuracies, attributed to the reduction of spurious marker associations arising from the pedigree.

Reference vs. Prediction Populations. One of the key components to utilizing genomic prediction of breeding values is to capitalize on the LD between markers and QTL. Many of the modeling methods and considerations have been outlined above, but also of importance is relationship and size of the training (i.e. reference) population to the population meant for prediction of breeding values (de Roos et al., 2009; Ibáñez-Escriche et al., 2009; Habier et al., 2010b; Kizilkaya et al., 2010; Saatchi et al., 2010; Toosi et al., 2010). For example, de Roos et al. (2009) simulated 2 divergent lines and used varying mixtures of those populations to create training and prediction populations. From this study, they concluded that as the relatedness of breeds, and therefore animals, in training populations decreased, the density of markers must also increase to maintain the LD between markers and QTL.

This conclusion was supported by Ibáñez-Escriche et al. (2009), as they evaluated the predictions of genomic breeding values using simulated crossbred populations to predict purebred performance. The use of breed-specific effects of SNP alleles did not yield any extra benefit over using across-breed effects of SNP genotypes unless the populations were very divergent and therefore had large breed effects. In regards to reference population size, the breed and relationship composition is of importance, as Ibáñez-Escriche et al. (2009) noted that the use of a 4-breed cross training set required 4 times as many records as training in a purebred line to retain the same accuracy because the effective population size is increasing. Overall, the use of across

breed or population predictions was most valuable when the populations were closely related, there were high density of markers, and the number of records in the training set was small when compared to breed-specific predictions (Ibáñez-Escriche et al., 2009).

Saatchi et al. (2010) studied, specifically, the effect of the training record size and relationship of training to prediction population, and concluded that as the number of records increased, the accuracy increased and generations of training animals closer to the prediction animals yielded higher accuracies. As Habier et al. (2007) noted, Saatchi et al. (2010) concluded that differences in accuracy based on training and prediction populations could be due to 1) the strength of relationship between the 2 populations, 2) the amount of recombination that can occur over time, and 3) the reduction of LD over generations.

Application to Study

The development of these Bayesian methodologies to handle genome-wide association and prediction studies must relate back to the original goal – to derive accurate breeding values for selection. Without the development of these procedures into usable tools for breeders of livestock, no benefits will be seen despite the optimistic expectation of faster, more cost effective genetic gain and reduction of generation interval. Application of genomic selection has already started in the dairy industry (Hayes et al., 2009; VanRaden and Sullivan, 2010; Rius-Vilarrasa et al., 2012), and investigation with beef breeds has started (Saatchi et al., 2011), where results from simulation studies may indicate that faster genetic gain is possible, while also controlling

the amount of inbreeding, due to selection at earlier ages (Daetwyler et al., 2007; Calus and Veerkamp, 2011; Ibáñez-Escriche and Blasco, 2011).

Many studies have used simulation methods with a few using actual animals from dairy (Verbyla et al., 2009; Habier et al., 2010b; Rius-Vilarrasa et al., 2012) and beef (Bolormaa et al., 2011; Saatchi et al., 2011) breeds. Furthermore, application of these methodologies to conduct association studies has begun, particularly in pigs (Onteru et al., 2011; 2012; Boddicker et al., 2012), while some simulation studies have been conducted using Bayesian methods for both QTL detection and prediction of breeding values (i.e., Veerkamp et al., 2010).

The population of *Bos indicus* x *Bos taurus* cattle currently being researched at Texas A&M University offers a unique opportunity to investigate the use of genomic selection methodologies, specifically with Bayesian methods, for various traits. In addition, the design of the family structure provides the ability to trace origin of alleles to the founding breeds and to evaluate the impact that has on modeling and prediction of breeding values. It is, therefore, of interest to utilize this population for analysis of genome-wide association and selection strategies using Bayesian inference.

CHAPTER III

MATERIALS AND METHODS

Cattle

The Texas A&M McGregor Genomics Cycle 1 population consists of 14 full-sibling F_2 families and 4 paternal half-sibling families. The F_2 families were founded by Nellore (N) grandsires and Angus (A) grand-dams. Five F_1 sires (identification numbers: 297J, 432H, 437J, 551G, and 2855) were mated to 13 F_1 dams to produce 15 full-sibling families by multiple ovulation and embryo transfer (ET), where two of the females were used with 2855 and another bull (i.e., used twice). Semen quality for one of the F_1 bulls (2855) was inadequate, so use of the bull was discontinued and the 2 progeny produced by that sire were excluded from all analyses, which left 13 full-sibling F_2 families available for this study. The remaining 4 F_1 sires were also mated to an additional population of F_1 and F_2 *Bos indicus* x *Bos taurus* cows to produce paternal half-sibling families through natural service. These calves were produced in multiple-sire pastures and required DNA testing to determine paternity. A total of 480 F_2 ET and 266 natural service calves were produced from 2003 to 2007.

The Texas A&M McGregor Genomics Cycle 2 and 3 populations were a continuation of crosses using N and A breeds. The Cycle 2 population consists of all possible combinations of reciprocal natural service F_2 calves produced from 2009 to 2011, including NA x NA, NA x AN, AN x AN, and AN x NA crosses of F_1 cattle, where pairs of letters indicate breed of sire and dam in both parents. The Cycle 3

population was produced at the same time as the Cycle 2 calves and consists of F_3 calves by natural service 1) whose sires were Cycle 1 F_2 ET bulls sired by 432H or 437J and whose dams were Cycle 1 F_2 ET females sired by 297J or 551G and 2) whose sires were Cycle 1 F_2 ET bulls sired by 297J or 551G and whose dams were Cycle 1 F_2 ET females sired by 432H or 437J. In total, 9 F_2 ET bulls produced calves in Cycle 3. There were 169 calves of the Cycle 2 and 3 matings produced from 2009 to 2011 and had records available. Only steers (n = 70) from Cycles 2 and 3 were genotyped.

Genotypes

Blood samples were previously collected on all live-born animals in the population, and DNA was extracted to use for genotyping. Cycle 1 animals, their parents and grandparents were genotyped using the Bovine SNP50 Version 1 assay (Illumina Inc., San Diego, CA) that produced genotypes for 54,001 SNP, whereas Cycle 2 and 3 steers were genotyped using the Bovine SNP50 Version 2 assay that produced genotypes for 54,609 SNP. Quality checks were completed for the combined dataset, and a marker was excluded if it had less than 90% of the animals successfully genotyped, minor allele frequency was less than 0.05, genotype frequencies significantly deviated from Hardy-Weinberg Equilibrium proportions or a combination of any of the previously stated criteria. After editing, 34,913 SNP were available for use.

All markers were mapped to UMD 3.1 assembly from the University of Maryland, released in May 2010 (*Bos taurus* 6.1 on NCBI, http://www.ncbi.nlm.nih.gov/ projects/mapview/map_search. cgi?taxid=9913). UMD 3.1 marker coordinates were

used for all association and prediction analyses in this study. All procedures involving animals were approved by the Texas A&M Institutional Care and Use Committee: AUP 2002-116, 2005-147, 2008-234 and 2011-291.

Traits

Overall Temperament at Weaning. All calves were scored at weaning for 5 subjective temperament scores including aggressiveness, nervousness, flightiness, gregariousness, and overall temperament by 4 evaluators, which were previously described by Boldt (2008) and Funkhouser (2008). Evaluators assigned scores independently on a 1 to 9 scale, where 1 was associated with the calmer, more docile animals, and 9 was associated with the wilder animals. Prior to scoring, calves were separated into groups of approximately 15 animals and placed in holding pens. In the Cycle 1 animals, two calves were separated out of the group into an alleyway approximately 25 meters long with 2 evaluators standing at either end of the alley. After a short period of time, one of the two calves in the alleyway was placed back into the holding group and the remaining calf was evaluated for all 5 traits before being turned into a separate holding pen. For Cycle 2 and 3 animals, only one calf was separated into the alleyway and was then given a brief period to calm down before being evaluated and released into the separate holding pen. Overall temperament at weaning was the analyzed trait in this study. Calves at weaning with both genotype and phenotype information for overall temperament at weaning for Cycles 1, 2, and 3, excluding bulls

from Cycle 1, were used (n = 772). Temperament scores were averaged across evaluators to create a single record per animal.

Warner-Bratzler Shear Force. After Cycle 1 calves were weaned, they were released to graze on pasture for approximately 130 d to grow. Once steers reached an average age of 11 to 13 mo, individual feed intake was evaluated using a Calan gate system (American Calan, Inc., Northwood, NH), described by Amen (2007). Steers were housed in a partially covered facility with pens of 4 animals and weighed every 28 d. Feed was offered *ad libitum* and consisted of 90% dry matter on average. Feed not eaten was collected every 7 d and weighed. Fresh feed was offered when refused feed built up to a large extent. A few steers never learned to eat from the Calan gate bunks and were placed in adjacent individual pens and fed the same ration.

After Cycle 2 and 3 calves were weaned, steers were transferred to the Texas A&M University Beef Systems Research Unit near College Station and placed on a high-forage growing diet using GrowSafe® feed intake observation units (GrowSafe Systems, Ltd, Airdrie, AB). Steers were kept in 4 pens of approximately 23 steers per pen and fed *ad libitum* approximately 6 mo before being transferred to Graham Land & Cattle Company in Gonzales County, Texas. Once at Graham, steers were placed on a typical finishing diet.

Cycle 1 steers were harvested, after approximately 140 d on feed (i.e., approximately 18 mo of age), in 6 groups over a 3 yr period at Texas A&M University Rosenthal Meat Science and Technology Center in College Station, described by Nicholson (2008). After the carcass was split, only the right side was electrically-

stimulated through a probe (Koch Britton Stimulator 350, Kansas City, MO), which was inserted into the neck muscles at 550 V for 3 sec, repeated 20 times with a 1 sec rest period between stimulations. After slaughter and stimulation (approximately 45 min postmortem) carcasses were placed in a cooler and allowed to chill for 48 h. Following chilling, the loin was removed from both sides of the carcass and used to cut 2.54 cm thick steaks from the most anterior part of the loin, vacuum-packaged, and held in a cooler at 2°C until 14 d postmortem before being frozen in a -10°C freezer.

Cycle 2 and 3 steers were harvested, after approximately 140 d on feed (i.e., approximately 18 mo of age), in a single group, per year, at Sam Kane Beef Processors in Corpus Christi, Texas. Both sides of the carcasses were electrically-stimulated with three stimulation bars for 27 sec each. The first bar supplied 150 V at 1.9 amp and the remaining two supplied 300 V at 3.0 amp. Carcasses were chilled at 0°C for 48 h postmortem. Following chilling, the loins were vacuum packaged and shipped to the Texas A&M University Rosenthal Meat Science and Technology Center under refrigeration. Six 2.54 cm steaks from the anterior end of the loin were cut and randomly assigned to 1, 7, or 14 d aging periods for Warner-Bratzler shear (WBS) force determination. Steaks were vacuum-packaged and stored at 1°C for the designated period of time. Steaks aged for 14 d were used in this study.

Before cooking, steaks were thawed (if frozen) in a 4°C cooler for 48 h. Grated non-stick electric grills (Hamilton Beach™ Indoor/Outdoor Grill, Southern Pines, NC) were used to cook the steaks and were pre-heated for 15 min so that they reached an approximate temperature of 177°C. Internal temperature of each steak was monitored

with a thermometer (Omega^{TM} HH501BT, Stamford, CT). Each steak was turned once after reaching an internal temperature of 35°C, removed once it reached an internal temperature of 70°C, and cooled approximately 4 h or until reaching room temperature. Following cooling, steaks were trimmed to remove visible connective tissue and to expose the muscle fiber orientation. For each steak, at least six 1.27 cm diameter cores were removed parallel to the orientation of the muscle fibers and sheared once perpendicular to the muscle fibers using a United Testing machine (United 5STM-500, Huntington Beach, CA) with a 11.3 kg load cell and a v-notch WBS force attachment. The peak force required to shear each core was recorded. As electrically-stimulated data is only available on Cycle 2 and 3 steers, only electrically-stimulated data from Cycle 1 steers will be used for this study. The average shear value, averaged across all shears per steak, for 14 d aged steaks from steers (n = 390) with both genotype and phenotype data were used in this study.

Statistical Analysis

For each trait, general linear or mixed model procedures of SAS (SAS Institute Inc., Cary, NC) were used to develop the final fixed effects model for each trait that would be used for the rest of the objectives in this study. These were formatted according to the software requirements, including recoding to identify correct nested effects (e.g. family nested within sire). Fixed effects investigated were sire, type of cross, family, sex, birth year-season combinations, pen nested in birth year-season combinations, and date of shear (averaged 25 steers per group), and other effects

depending on the trait (see Table 2 for number of levels per fixed effects factors). Alternative association and genetic merit prediction models were investigated using breed-of-origin SNP genotypes in addition to or without original genotypes, for overall temperament at weaning. Because of this, fixed effects accounting for similar family structure (i.e., sire, family, or both) were excluded to avoid confounding effects.

Independently of sire and family nested within sire, type of cross was evaluated if either sire or family nested within sire were not significant, which was the case for Warner-Bratzler shear force. This effect consists of levels based on sire breed combination for each cycle (e.g., Nellore-Angus F_1 or F_2) paired with dam breed combination (Table 3). For both traits, animals with data for fixed effects in the final model, phenotypes and genotypes were used in the study (n = 769 or 389 for overall temperament at weaning or WBS force, respectively).

Table 2. Number of levels per fixed effect factor evaluated for overall temperament at weaning and Warner-Bratzler shear force[1]

	Overall temperament at weaning		Warner-Bratzler shear force[2]	
Fixed Effect Factor[3]	Cycle 1	All cycles	Cycle 1	All cycles
Sire	4	16	4	16
Family	17	30	13	26
BYS	9	10	9	10
Pen(BYS)	38	43	-	-
Sex	2	2	-	-
Type of cross	-	-	10	12
Date of shear	-	-	13	16

[1]Cells with "-" indicate the effect was not evaluated.
[2]Trait only included steers and therefore sex was not evaluated.
[3]Fixed effects evaluated but may not be included in the final model. Family and Family nested within sire are identical and therefore only Family is listed. BYS is birth year-season combinations and Pen(BYS) is pen nested within birth year-season combinations. Type of cross refers to the sire and dam breed combination (see Table 3). Date of shear is a contemporary group effect that groups steers by the date in which the steaks were evaluated for Warner-Bratzler shear force.

Table 3. Levels and number of animals per level for type of cross fixed effect

Type of cross[1]	Cycle 1	All cycles
NA_ABBA	11	11
NA_BA	29	29
NA_BA2	7	7
NA_BH	18	18
NA_BH2	3	3
NA_BHHB	4	4
NA_HB	12	12
NA_HB2	10	10
NA_HBBH	12	12
NA_NAF2	214	224
AN_NAF2	0	10
NA_NAF3	0	49
Total	320	389

[1]The "_" separates the progeny's sire and dam's breed composition, where pairs of letters represent the sire and dam breed or cross, respectively. Breeds include Nellore (N), Angus (A), Hereford (H), and Brahman (B).

Comparison Criteria for Breeding Value Prediction

Studies have shown that Pearson correlation coefficients (r) between the true breeding value (BV) or deregressed estimated BV and the genomic estimated breeding value (GEBV) can be used to assess accuracy (e.g., Meuwissen et al., 2001; Saatchi et al., 2011). When raw phenotypes are used, this correlation underestimates the true accuracy (Saatchi et al., 2012). Therefore, accuracy for this study was calculated following the standardization method proposed by Saatchi et al. (2011), and takes the form of:

$$Accuracy = \hat{\rho}_{g\hat{g}} = \frac{\sigma_{P,EBV}}{\sigma_g \sigma_{\hat{g}}},$$

where $\sigma_{P,EBV}$ is the covariance of the phenotype or trait with the estimated breeding values, $\sigma_{\hat{g}}$ is the standard deviation of the estimated additive genetic effect from the sample population, and σ_g is the standard deviation of the additive genetic effects from the entire population available (i.e., animals with phenotypic and genotypic information), calculated as:

$$\sigma_g = \sqrt{\sigma_g^2} = \sqrt{h^2\sigma_P^2},$$

where h^2 is the heritability of the trait and σ_P^2 is the phenotypic variance calculated from the animals with available genotypic data. Reliability ($\hat{\rho}_{g\hat{g}}^2$) is the square of this correlation, which will not be reported in study. The simple linear regression coefficient ($\hat{\beta}_{y,x}$), which gives an indication of biasness (Saatchi et al., 2013) was also used to compare models and evaluate performance across validation groups. The simple linear regression coefficient ($\hat{\beta}_{y,x}$) is the regression of phenotype on EBV and takes the form of:

$$\hat{\beta}_{y,x} = \frac{\sigma_{y,x}}{\sigma_x^2} = \frac{\sigma_{EBV,P}}{\sigma_{EBV}^2},$$

where $\sigma_{EBV,P}$ is the covariance of EBV and phenotype and σ_{EBV}^2 is the variance of the EBV. This parameter should be 1 if unbiased (Saatchi et al., 2011; 2013).

Because genetic merit is being considered, ranking of individuals based on their EBV or GEBV is of importance, especially if re-ranking occurs due to model parameters. Spearman Rank correlation coefficients were assessed to understand the overall magnitude of re-ranking of estimated breeding values that occurred between any two analyses when comparing model types (e.g., BayesC versus BayesB), where animals were ranked from top to bottom depending on the desirability of that particular trait (e.g., calmer individuals would be ranked higher than wild or crazy individuals). In particular, the percentage and number of individuals whose breeding values changed quartiles between any two analyses were assessed, where it was of interest to identify the number of individuals whose breeding values changed up to 3 quartiles (which could occur in only half of the population).

Bayesian Methods Employed

Prediction of breeding values using genomic information is a two part process using the Bayesian framework of 1) training the markers to the data to calculate substitution effects using both phenotypes and genotypes and 2) using genotypes to calculate GEBV based on training results. The models employed for this study take the form of a linear mixed model to calculate fixed and random effects such as:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\alpha + e,$$

where $\boldsymbol{y}$ is an n x 1 vector of phenotypic values (i.e., observations), $\boldsymbol{X}$ is an n x p incidence matrix that relates the n x 1 non-genetic fixed effects (β) to y, $\boldsymbol{Z}$ is an n x k matrix of genotype covariates (coded as 0, 1, or 2) for k SNP markers, α is a k x 1 vector of marker effects, and e is a n x 1 vector of residuals (see Fernando and Garrick, 2013). In all cases, the prior distributions for these Bayesian regressions must be specified for β, α, and e, where β utilizes a flat prior and requires information on the residual variance (σ_e^2). For residuals, a normal distribution with a null mean and covariance matrix $\boldsymbol{R}\sigma_e^2$ is used, where $\boldsymbol{R}$ is an identity matrix and therefore assumes identical variance and no covariance between observations. The residual variance (σ_e^2) is treated as an unknown parameter and assumes a scaled inverted chi-square prior. Further information regarding prior specifications for both β and σ_e^2 are described in Fernando and Garrick (2013). The Bayesian methods employed in this study differ primarily in their assumptions related to the prior for the variance component of the marker effects (α).

For the BayesB method (Meuwissen et al., 2001), each marker is assumed to have a locus specific variance σ_i^2 that comes from a scaled inverted chi-square prior using a scale parameter S_α^2 and v_α degrees of freedom (Table 1) and fits a mixture model that assumes some known fraction of markers (π) has zero effects. BayesC method (Kizilkaya et al., 2010; Habier et al., 2011) utilizes a mixture model as well, but markers included in the model share a constant genetic variance σ_α^2 (Table 1) with a scaled inverted chi-square prior.

For this study, MCMC methods with 51,000 iterations were implemented to provide posterior mean estimates of marker effects and variances with the first 1,000 of

those iterations discarded to avoid bias due to starting values. Parameter estimates were produced every iteration and marker window variances were produced every tenth iteration (i.e., output frequency = 10). Estimates of genetic ($\hat{\sigma}_g^2$) and residual ($\hat{\sigma}_e^2$) variances used for starting values were obtained as follows:

$$\hat{\sigma}_g^2 = h^2\sigma_P^2 \text{ and } \hat{\sigma}_e^2 = (1 - h^2)\sigma_P^2,$$

where h^2 is the heritability of the trait as identified from the literature or preliminary analyses with the present population and σ_P^2 is the phenotypic variance estimated from the data.

Both BayesB and BayesC require that the parameter π, a proportion of markers not contributing to the trait of interest, is known. This is rarely the case, however. BayesCπ procedures, which samples π from a uniform prior, were implemented to estimate π for further analyses. Seed values of 0.1 and 0.9 were used to ensure that convergence was reached. In some cases, the genetic architecture of the trait and the number of observations available hinder the ability of BayesCπ to converge to an appropriate estimate (preliminary runs, data not reported), which occurred for both traits in this study.

Because of this, the genomic heritability estimate (h_g^2, a ratio of the posterior means of the genetic and phenotypic variances) was found for each trait using BayesC procedures with π = 0 (i.e., all markers included in the analysis). From there, π was increased in intervals while keeping other parameters constant until a slight drop (0.01 to

0.02) in genomic heritability was observed, which determined $\hat{\pi}$. All analyses were performed using GenSel software (Fernando and Garrick, 2009).

Objective 1: Association and Ontology Analyses

Using $\hat{\pi}$, BayesB and BayesC procedures were employed with the "windowBV" option in GenSel to calculate association results for overall temperament at weaning and WBS force. This option separates markers into 1 Mb chromosomal segments based on the marker map provided (UMD 3.1, summary information is provided in Table 4). Association for a 1 Mb window was determined by the posterior probability of association for that window (PPA_w). As results accumulate over N independent experiments for each genomic interval, the proportion of true associations converges to the PPA_w and the proportion of false positives (PFP) converges to 1 - PPA_w (Fernando et al., 2004; Fernando and Garrick, 2013). In the case of using $\hat{\pi}$, it was of interest to determine if a window accounted for greater than 0 percent of the genetic variance a large number of times and, therefore, resulted in PPA_w in the top 25^{th} percentile (i.e., $PPA_w > 0.75$). Windows that met this criterion were identified. Association plots were generated using R software (R Development Core Team, 2008) with the script provided in Appendix B.

Based on preliminary association results for overall temperament at weaning and WBS force, association results using BayesC procedures with $\pi = 0$ were also evaluated with the same threshold value for significance (0.75). Evaluating association results in this scenario would imply that the trait's architecture may follow an infinitesimal model

Table 4. Average and range of the number of markers per 1 Megabase (Mb) window per chromosome

Chromosome	Average number of markers per Mb window	Range of the number of markers per Mb window
Unknown[1]	9.91	8 - 10
1	13.83	2 - 22
2	13.40	4 - 28
3	13.27	4 - 22
4	13.41	3 - 22
5	11.43	1 - 23
6	14.15	2 - 22
7	13.63	1 - 23
8	13.87	6 - 23
9	12.35	2 - 21
10	13.99	4 - 24
11	14.09	4 - 24
12	12.56	3 - 22
13	14.28	5 - 25
14	13.90	4 - 27
15	13.37	1 - 23
16	13.79	2 - 21
17	14.11	2 - 23
18	13.36	1 - 22
19	13.92	1 - 23
20	14.56	5 - 22
21	13.19	1 - 24
22	14.19	6 - 22
23	13.75	2 - 24
24	13.92	5 - 28
25	14.86	5 - 25
26	13.94	6 - 22
27	14.22	5 - 21
28	13.28	4 - 23
29	13.54	2 - 23
X	3.65	1 - 19

[1]Markers listed under this chromosome were unassigned to a chromosome based on *Bos taurus* UMD3.1 build.

and, therefore, many more windows are expected to be associated. The PPA_w in this case assumes a different meaning. The posterior probability of association when $\pi = 0$ (identified as PPA_{w2}) would consider if a window accounted for more than the average genetic variance of all windows (i.e., the genetic variance is greater than average expected, calculated by dividing the genetic variance by the number of windows). Because a larger number of windows are expected to be identified in this scenario, additional steps were taken to evaluate association results using a systematic approach.

To do this, markers were first mapped to the closest gene by developing an R script (see Appendix B). This script first accesses the feature information for the specific species' build on the NCBI server (e.g. *Bos taurus* Build 6.1), selects the assembly and features the user prefers (i.e., gene, pseudogene, RNA, etc), then finds the genes closest to each marker supplied by the user based on the map information (UMD3.1 build).

From this, a list of genes for each trait was identified based on markers that fell in regions meeting $PPA_{w2} > 0.75$ criteria. Each trait's associated gene list was used for enrichment analysis, the identification of gene ontology (GO) terms that are significantly overrepresented in a given set of genes, using GOrilla online software (Eden et al., 2009). *Homo sapiens* gene information was used (as *Bos taurus* is not currently available) along with a background list of unique genes (n = 12,566) that could have been identified by any of the markers available for use (based on the output from the R script). Gene ontology term results were reported if found to be significant when related to biological processes, cellular components, and molecular functions. The *P*-value threshold was set to 10^{-3}, but this criterion did not correct for testing across multiple

gene ontology terms. GOrilla also reports a false discovery rate (FDR) q-value as means to correct for multiple testing. The equation used to calculate this value was:

$$FDR\ q-value = \frac{p*N}{i},$$

where *p* is the *P*-value associated with the gene ontology term in the analysis, *N* is the total number of GO terms, and *i* is the rank of the GO term using *P*-values for the analysis. The FDR q-value is based on Benjamini and Hochberg (1995) calculation of FDR. Separate FDR q-values are calculated based on the number of terms in biological processes (n = 10,150), cellular components (n = 1,165), or molecular functions (n = 3,146). In addition, GOrilla reports enrichment scores calculated as:

$$Enrichment = \frac{\left(\frac{b}{n}\right)}{\left(\frac{B}{N}\right)},$$

where *N* is the total number of genes from the background list recognized and not duplicated, *B* is the total number of genes associated with a particular GO term being evaluated, *n* is the number of genes from the target set supplied by the user that is recognized and not duplicated, and *b* is the number of genes in *n* that are associated with the particular GO term being evaluated. The enrichment score is a ratio based on the proportion of associated genes found in the target set compared to the proportion of

associated genes for that ontology term out of the total number of genes supplied in the background set.

Objective 2: Comparison of Models on Impact of Breeding Value Prediction

It was of interest to compare different methods of Bayesian modeling along with the traditional animal model in order to determine the effectiveness of each method. For each trait, a traditional animal model (i.e., traditional BLUP) was employed that incorporated fixed effects developed in SAS, random animal effects using pedigree information, and random residuals with the ASReml software (Gilmour et al., 2009) for both traits. BayesB and BayesC procedures were employed with $\hat{\pi}$ in addition to BayesC with $\pi = 0$ (as previously described). All animals available were included for the training phase. GenSel automatically generates GEBV for all animals in the training set; therefore, no additional steps were needed at this time. Results from these analyses were compared based on the comparison criteria previously outlined.

Objective 3: Inclusion and Impact of Breed-of-Origin Genotype on Breeding Value Prediction

The population structure provides a unique opportunity to assess the origin of each individual's alleles and the potential impact on performance of predicting genetic merit using a Bayesian model. For the three generation pedigree available of the F_2 ET cattle, haplotype phase was determined using fastPHASE software (Scheet and Stephens, 2006) for every 1 Mb, sliding in overlapping ½ Mb windows (excluding the

chromosome ends). This generated 4,783 output files across the 30 chromosomes (29 autosomes and X chromosome). These output files were parsed using an algorithm developed in Perl (http://www.perl.org/) and genotypes based on the breed-of-origin for each allele (0, 1, or 2 based on the number of N alleles and referred to as breed-of-origin genotypes) were assigned by following the phased haplotypes through the three generation pedigree including grandparents, F_1 parents, and F_2 offspring (C. A. Gill (Texas A&M University, College Station, TX), personal communication). Often, the sequencing depth of the X chromosome is about half that of autosomes, resulting in lower SNP detection rate and potential genotyping errors, based on the sex of the animal, are possible (Zhan et al., 2011). To be conservative with the use of breed-of-origin genotypes, which are based on assembly and phasing techniques, genotypes available on the X chromosome were not used to avoid any potential false associations that may arise due to these procedures.

Assembly errors among the 29 autosomes and recombination events did leave various regions among the animals without full breed-of-origin genotypes. To utilize these genotypes, missing data was addressed in 3 ways by developing an algorithm in R software (R Development Core Team, 2008; Appendix B). First, if a region with missing data had both ends with the same genotype (e.g., a breed-of-origin genotype of AA (coded as 0) at one end and AA (coded as 0) at the other end), then the missing data was inferred to be the same genotype as the respective ends (e.g., AA (coded as 0)) and replaced with the corresponding genotype code. Second, if a region with missing data had ends that differed (e.g., AA at one end but heterozygous (NA or AN, coded as 1) at

the other end), then missing data for that region was replaced with the family average for those markers. Family average genotypes were calculated by taking all animals within that family with genotypes available, calculating the average genotype, and rounding to the respective whole number to keep coding consistent (i.e., 0, 1, or 2). Third, there were regions within a particular family that had missing data even after calculating the family average. In this case, missing data was replaced with the value of a heterozygote (i.e., 1).

After formatting, breed-of-origin genotypes were used to run a genomic prediction analysis for overall temperament at weaning with or without the original nucleotide-based genotypes using $\hat{\pi}$ with a Bayesian model (as determined by results in Objective 2). Fixed effects related to family structure (i.e., sire, family, or both) were excluded.

Objective 4: Impact of Dataset Partitioning

Relationship of the training population to the prediction or validation population can have a direct impact on results (de Roos et al., 2009; Ibáñez-Escriche et al., 2009; Habier et al., 2010b; Kizilkaya et al., 2010; Saatchi et al., 2010; Toosi et al., 2010). Overall temperament at weaning was used to determine the effect of data partitioning in this population, because of the larger number of records for this trait (n = 769). A Bayesian model for this trait (as determined by results in Objective 2) was used to generate GEBV for training and validation groups. Breeding values generated using Bayesian methods were compared to traditional EBV generated using ASReml software. Phenotype files used for training procedures in GenSel were used to generate the EBV in

ASReml, after reformatting, which would result in animals in the validation population to have EBV predicted without phenotype records. To predict breeding values in GenSel for validation animals, the "Predict" procedures (Fernando and Garrick, 2009) were run, which sums across posterior means of random marker effects for each individual to formulate the GEBV.

Initial analyses were conducted using only Cycle 1 progeny in the validation set, where training was conducted (1) once using all animals in training, (2) 4 times using the progeny of all but one sire, (3) once using only embryo transfer F_2 progeny, and (4) once by random assignment (Table 5). Additional analyses were run to increase sample numbers that included Cycles 2 and 3. Training was conducted (1) 4 times using the progeny of all but one sire or grandsire, (2) once using only F_2 progeny, (3) once by random assignment, and (4) once using only Cycle 1 progeny (Table 5) for these analyses. In the case of random assignment, the number of animals to be included in the validation population was determined by taking the average of the 5 previously listed scenarios (e.g., 1 and 2 in both cases). Whether only Cycle 1 progeny or progeny from all cycles animals where selected to be in the validation population, the sample sizes did not change much, but relationships between training and validation groups had the potential to change.

Table 5. Sample numbers of training and validation populations for overall temperament at weaning

	Population	
Analysis[1]	Training	Validation
None	769	0
Sire is 297J	595	174
Sire is 432H	583	186
Sire is 437J	577	192
Sire is 551G	612	157
Natural service offspring	441	328
Random (cycle 1 only)	592	177
Sire or grandsire is 297J	588	181
Sire or grandsire is 432H	576	193
Sire or grandsire is 437J	566	203
Sire or grandsire is 551G	587	182
Non-F2 offspring	461	308
Random (all cycles)	579	190
Cycles 2 and 3	699	70
Average	588	182

[1]Analysis refers to the progeny selected to be in the validation population. The remaining animals were used in the training phase.

CHAPTER IV

RESULTS AND DISCUSSION

Statistical Models Developed

Overall Temperament at Weaning. Modeling of fixed and random effects for overall temperament at weaning were previously described by Funkhouser (2008) for Cycle 1 animals, which included 1) fixed effects of sire, family nested within sire, birth year-season combinations, sex, pen within birth year-season combinations, evaluator within birth year-season combinations, a sex by sire interaction, and 2) random effects of recipient disposition within birth year-season.

The data for this study incorporate genotyped individuals from Cycles 1, 2 and 3, and the Bayesian genetic prediction software does not accommodate random effects (besides markers) to be included in the model. Because of this, a reduced model was evaluated for use in this study. The average overall temperament score for each animal was used to evaluate fixed effects of sire, family nested within sire, sex, birth year-season combinations, pen within birth year-season, and sex by sire interactions. Effects were investigated in preliminary analyses with the GLM procedures of SAS (SAS Institute Inc., Cary, N. C.).

The interaction of sex by sire was not found to be significant when considering only Cycle 1 animals or combination of Cycle 1, 2, and 3 animals ($P > 0.99$). Significance of the sex by sire interaction was driven by the inclusion of bulls in Funkhouser (2008), but only females and steers were used in this study, and therefore

the interaction was not expected to be significant. Sire was found to be significant; however, the design of the sire effect for the natural service half-sibling, Cycle 2, and Cycle 3 animals meant that sire was equivalent to family nested within sire. Therefore this effect was not included to avoid bias due to redundancy. Significance levels for fixed effects in the final model are reported in Table 6 for both scenarios. There was concern on including family nested within sire as a fixed effect for prediction of genetic merit as it may remove some genetic variation among families. Preliminary results indicated genotype frequency differences between families, which could result in stratification and possible false associations (e.g., Lander and Schork, 1994; Marchini et al., 2004, Janss et al., 2012). To avoid this, family nested within sire was included for all analysis, but analyses may be run in the future without the effect to determine impact on prediction of genetic merit.

Warner-Bratzler Shear Force. Fixed effects of sire, family nested within sire, birth year-season combinations, and date of shear were evaluated. Sex was not evaluated as all animals in these analyses from Cycles 1, 2, and 3 were steers. When considering

Table 6. Probability values for *F* ratios of fixed effects evaluated for overall temperament at weaning

Fixed Effect	Cycle 1[1]	All cycles[2]
Sex	0.004	0.004
Family(sire)	< 0.001	< 0.001
Birth year-season	< 0.001	< 0.001
Pen(birth year-season)	0.003	0.003

[1]Cycle 1 consisted of 702 animals, where 699 had information available for all fixed effects evaluated.
[2]All cycles refer to Cycles 1, 2, and 3, which consisted of 772 individuals with 769 having available information for all fixed effects evaluated.

only Cycle 1 animals and when considering the combination of all cycles, fixed effects of sire, family nested within sire, and birth year-season combinations were not contributing to the variation in WBS force ($P > 0.25$). Because sire and family nested within sire were not significant, type of cross was evaluated and was significant. In all cases, date of shear was significant. The final model included type of cross and date of shear as fixed effects. *P*-values for fixed effects are reported in Table 7.

Table 7. Probability values for *F* ratios of fixed effects evaluated for Warner-Bratzler shear force

Fixed effect	Cycle 1[1]	All cycles[2]
Type of cross[3]	0.042	0.020
Date of shear	< 0.001	< 0.001

[1]Cycle 1 consisted of 320 animals, where all 320 steers had information available for all fixed effects evaluated.
[2]Cycles 1, 2, and 3 steers (n = 389).
[3]See Table 3 for levels of this effect.

Trait Correlations. Previous significant and positive correlations between a measure of temperament and tenderness have been reported (e.g., King et al., 2006; Behrends et al., 2009; Hall et al., 2011). Common measurement of temperament uses exit velocity (Burrow et al., 1988), which is the speed at which an animal exits the working chute. The premise is that those animals that are calmer will exit the chute at a slower speed than those that are more excitable or wild. Times are captured by two infrared eyes (approximately 1.83 m apart) that connected to the timing unit. The time is started after the animal crosses the first infrared laser and stopped once the animal crosses the second laser. In addition, subjective evaluation methods have been used

including chute scores and pen scores (Grandin, 1993a; Grandin, 1993b), both on a 1 to 5 scale with 1 being calm or docile and 5 being excitable or wild.

Correlations between the two traits in this study were also assessed to determine any influence that temperament may have on tenderness measured through WBS force. All 389 steers that had WBS force data also had temperament data available. The Pearson's correlation (r) between overall temperament at weaning and WBS force for these 389 steers was 0.04 ($P = 0.41$). Spearman's Rank Correlation coefficient was calculated, but was not significant ($\rho = 0.09$, $P = 0.76$). The results from these correlations were in agreement with previous reports on a subset of these animals (see Nicholson, 2008), but do not concur with other studies on temperament and tenderness correlations (e.g., King et al., 2006; Behrends et al., 2009; Hall et al., 2011).

Objective 1: Association and Ontology Analyses

Start values for the additive genetic and residual variance were calculated using approximate heritabilities of 0.23 for overall temperament at weaning (based on preliminary runs using BayesC, $\pi = 0$, all animals included in the analysis), and 0.3 for WBS force (based on literature reports). O'Conner et al. (1997) estimated heritability on *Bos indicus* crossbred and composite cattle for electrically stimulated steaks aged 14 days. Their estimate of heritability was 0.27 ± 0.17. Estimates of heritability have ranged from 0.14 (Riley et al., 2003; Brahman cattle) to 0.53 (Shackleford et al., 1994; population of *Bos taurus* and Nellore cattle). Therefore heritability was approximated to be 0.3 to calculate start values for WBSF, which would give closer approximations than

the default start values provided in GenSel. Fixed effects used in association analyses included sex, family nested within sire, birth year-season combinations, and pen nested within birth year-season combinations for overall temperament at weaning and type of cross and date of shear for WBS force.

The posterior distribution for π from BayesCπ was not distinctly peaked, so it did not provide a precise estimate of $\hat{\pi}$ as previously discussed (see Appendix Figure A-1), but $\hat{\pi}$ was estimated to be 0.997 for overall temperament at weaning and 0.995 for WBS force determined by starting with $\pi = 0$ and incrementing its value until a slight drop was observed (D. J. Garrick (Iowa State University, Ames, IA), personal communication). Density plots of posterior estimates of genomic heritability using $\hat{\pi}$ are presented in Appendix Figure A-2. No window had a PPA_w greater than 0.75 when using $\hat{\pi}$ for either trait (Figure 1). Maximum PPA_w for overall temperament at weaning (0.51 or 0.39 for BayesB or BayesC, respectively) was on BTA 29 (3 Mb), but would not indicate association as PPA_w would equate to a PFP value of 0.49 and 0.61, respectively. The maximum PPA_w (0.19 or 0.17 for BayesB or BayesC, respectively) for WBS force was on BTA 29 (41 Mb) for BayesB and BTA 14 (10 Mb) for BayesC.

Regions with $PPA_w > 0.10$ are reported in Appendix Table A-1. A similar association profile for WBS force has previously been reported (McClure et al., 2012) including the well-known calpain 1, (mu/I) large subunit (BTA 29, 44 Mb), although calpastatin (BTA 7, 98 Mb) did not have $PPA_w > 0.10$ for this study. Previous studies on this population have resulted in peaks similar to Figure 1 for overall temperament at weaning as well, but not necessarily in the same Mb region (Wegenhoft, 2005; Boldt,

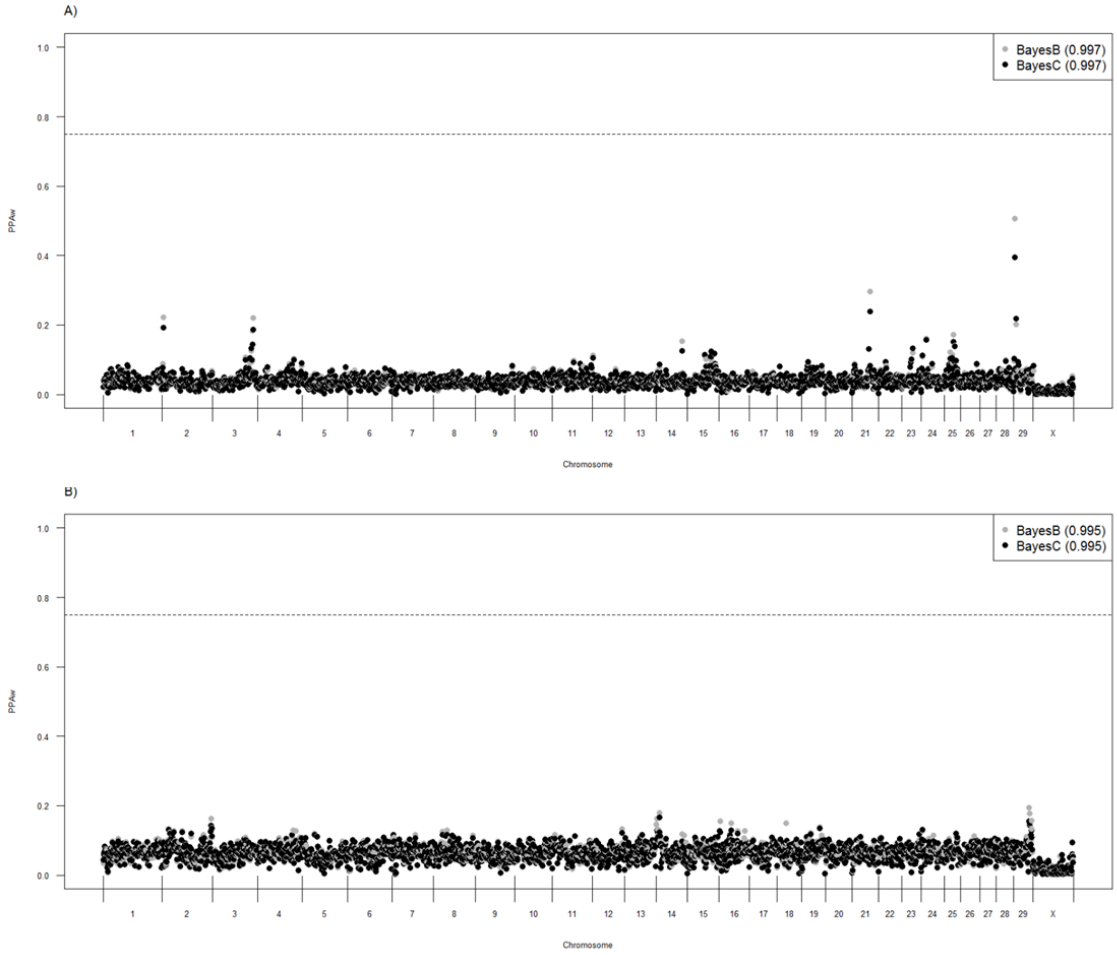

Figure 1. Posterior probabilities of association for windows (PPA_w) for all chromosomes on overall temperament at weaning (A) and Warner-Bratzler shear force (B) using BayesB and BayesC methods with the respective $\hat{\pi}$. The dotted line indicates the threshold for association ($PPA_w > 0.75$).

2008). Similarity in association profiles could indicate that the sample size is not large enough to find association with particular regions although the regions may be real QTL.

In addition, $\hat{\pi}$ values represent the best fit based on the available data and although providing best prediction of genomic merit, the parameter may not be the best

for QTL mapping. Calus et al. (2009) discussed the different aims between generating GEBV and QTL mapping using Bayesian output, where the aim of generating GEBV is to maximize the amount of QTL variance captured by the markers (or haplotypes in their study) and the aim of QTL mapping is to maximize the contrast in explained variance by the marker intervals. Although their paper focused on haplotype size differences, the principal difference between these two aims can still be applied in this study.

To investigate the possibility that $\hat{\pi}$ did not maximize the contrast in explained variance between markers for QTL mapping, output from BayesC with $\pi = 0$ was considered. Overall temperament at weaning had 37 windows and WBS force had 147 windows with $PPA_{w2} > 0.75$ (Figure 2, Appendix Table A-2). Of these windows, 38 and 1 windows were also identified as having $PPA_w > 0.1$ for WBS force and overall temperament at weaning when using $\hat{\pi}$, respectively. Using association output when $\pi = 0$ results in graphs atypical of traditional Manhattan style plots (see Figure 2), primarily due to chromosomes looking "noisier". This is most likely due to sampling techniques with posterior estimates and the posterior mean that are being produced over 50,000 iterations. Association results from analyses using $\pi = 0$, however, could still be useful in understanding biological terms and pathways influencing either trait.

Interestingly, every window identified for overall temperament at weaning (n = 37) was also identified for WBS force. Of these windows, however, only 2 had $PPA_w > 0.1$ when using $\hat{\pi}$ for both traits and 22 were identified by WBS force, but not overall temperament at weaning, as having $PPA_w > 0.1$ when using $\hat{\pi}$. King et al. (2006), Behrends et al. (2009), and Hall et al. (2011) reported positive correlations of

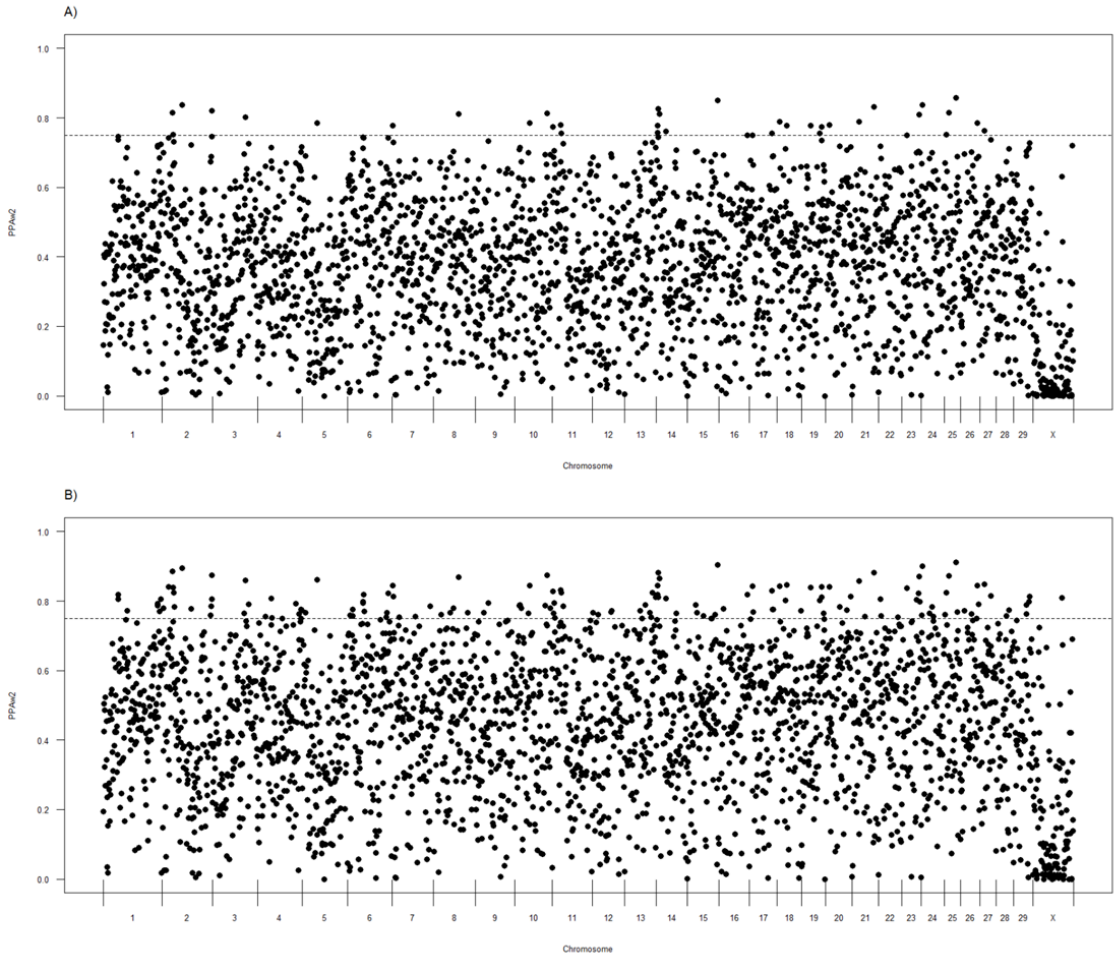

Figure 2. Posterior probabilities of association for windows (PPA_{w2}) for all chromosomes on overall temperament at weaning (A) and Warner-Bratzler shear force (B) using BayesC with $\pi = 0$. The dotted line indicates the threshold for association ($PPA_{w2} > 0.75$).

temperament with tenderness measures, although their measure of temperament differs from this study. No correlation was detected for the two traits in this study, which could be influenced by sample size, but it is possible that some of these regions contain genes or causal mutations influencing both traits are in the same chromosomal regions or have pleiotropic effects. This is most likely a small proportion, however, and identification of

similar regions could be due to other, non-biological factors as well (e.g., unaccounted for stratification within the data).

Enrichment analysis for overall temperament at weaning was conducted using 172 genes (located in associated windows when $\pi = 0$) with a background list of 7,225 genes recognized and not duplicated based on *Homo sapiens* information (see Table 8 for detailed description). Regions identified by overall temperament at weaning in this study have not previously been reported through the Cattle QTL database (Hu et al., 2013) or studies using subsets of this population (Wegenhoft, 2005; Boldt, 2008). Using genes identified by these markers revealed that ontological categories of biological processes, molecular functions, and cellular components only had significant *P*-values related to sodium ion transport (Table 8). In some cases, however, correction for multiple testing resulted in the term not being significant (FDR q-value > 0.05; Table 8).

Terms reported in Table 8 were investigated to understand the biological roles they could be identifying for overall temperament at weaning. Within the body, potassium and sodium ions help to regulate the membrane potential, the potential energy that comes from the charge separation across the membrane. In the nervous system, this is particularly important as voltage-gated ion channels for both potassium and sodium maintain the desired polarization in the resting state so that the axon membrane is ready for an impulse to occur (Sherwood, 2010). Briefly, a nerve impulse is the result of a series of action potentials, defined as the results once the charge difference across the membrane has reached the threshold value. In the case of nerve cells surrounded by the myelin sheath, action potentials only occur at the nodes of Ranvier (non-myelinated

Table 8. Gene ontology results for genes located in associated regions using BayesC with $\pi = 0$

Gene Ontology Term	Type[1]	Description[2]	P-value[3]	FDR q-value[4]	Enrichment (N, B, n, b)[5]	Trait[6]
Negative regulation of extracellular matrix disassembly (GO:0010716)	BP	Any process that decreases the rate, frequency or extent of extracellular matrix disassembly, the process that results in the breakdown of the extracellular matrix.	7.24E-04	1	11.12 (7462, 3, 671, 3)	WBSF
Sodium ion transport (GO:0006814)	BP	The directed movement of sodium ions (Na+) into, out of or within a cell, or between cells, by means of some agent such as a transporter or pore.	8.93E-04	1	5.25 (7225, 48, 172, 6)	OTW
Extracellular Region (GO:0005576)	CC	The space external to the outermost structure of a cell.	2.43E-06	0.003	1.71 (7462, 467, 671, 72)	WBSF
Sodium channel complex (GO:0034706)	CC	An ion channel complex through which sodium ions pass.	1.32E-04	0.075	14.00 (7225, 12, 172, 4)	OTW
T-tubule (GO:0030315)	CC	Invagination of the plasma membrane of a muscle cell that extends inward from the cell surface around each myofibril. The ends of T-tubules make contact with the sarcoplasmic reticulum membrane.	8.86E-04	0.505	4.10 (7462, 19, 671, 7)	WBSF
Voltage-gated sodium channel complex (GO:0001518)	CC	A sodium channel in a cell membrane whose opening is governed by the membrane potential.	9.95E-04	0.378	5.56 (7462, 10, 671, 5)	WBSF
			5.82E-05	0.066	16.80 (7225, 10, 172, 4)	OTW
Peptidase activity, acting on L-amino acid peptides (GO:0070011)	MF	Catalysis of the hydrolysis of peptide bonds formed between L-amino acids.	7.34E-04	0.525	1.69 (7462, 257, 671, 39)	WBSF

Table 8. Continued

Gene Ontology Term	Type[1]	Description[2]	P-value[3]	FDR q-value[4]	Enrichment (N, B, n, b)[5]	Trait[6]
Serine hydrolase activity (GO:0017171)	MF	Catalysis of the hydrolysis of a substrate[7]	2.93E-04	0.280	2.69 (7462, 62, 671, 15)	WBSF
Serine-type peptidase activity (GO:0008236)	MF	Catalysis of the hydrolysis of peptide bonds in a polypeptide chain[7]	2.42E-04	0.693	2.73 (7462, 49, 671, 13)	WBSF
Serine-type endopeptidase activity (GO:0004252)	MF	Catalysis of the hydrolysis of internal, alpha-peptide bonds in a polypeptide chain[7]	2.79E-04	0.399	2.95 (7462, 49, 671, 13)	WBSF
Voltage-gated sodium channel activity (GO:0005248)	MF	Catalysis of the transmembrane transfer of a sodium ion by a voltage-gated channel (a channel whose open state is dependent on the voltage across the membrane)	8.98E-05	0.254	15.27 (7225, 11, 172, 4)	OTW

[1]Type refers to biological process (BP), cellular component (CC), or molecular function (MF) in which the gene ontology (GO) term is classified as.
[2]Descriptions were accessed using identification numbers through AmiGO (amigo.geneontology.org)
[3]P-value was provided by GOrilla software (Eden et al., 2009) and computed based on the minimum hypergeometric (mHG) tail (not corrected for multiple testing).
[4]FDR q-value was provided by GOrilla software and is a corrected value of the p-value for multiple testing using Benjamini and Hochberg (1995) method, where for the ith term the FDR q-value = (p-value * number of GO terms)/i.
[5]Enrichment (N, B, n, b) was calculated by GOrilla software and is the N total number of genes recognized, non-duplicated, and associated with any GO term using *Homo sapiens* as the background species as *Bos taurus* is not available, B is the total number of genes associated with a given GO term, n is the total number of genes in the user defined list that are recognized and non-duplicated, and b is the number of genes from the user defined list that is associated with a given GO term. Enrichment = (b / n) / (B / N).
[6]OTW = overall temperament at weaning and WBSF = Warner-Bratzler shear force.
[7]These activities are by a catalytic mechanism that involves a catalytic triad consisting of a serine nucleophile that is activated by a proton relay involving an acidic residue (e.g. aspartate or glutamate) and a basic residue (usually histidine).

regions). This allows for a quick and fast transmission of the impulse through the nerve cell utilizing sodium ion transport and voltage-gated channels to the synapse connecting two nerve cells or the destination (usually termed the effector).

On a larger scale, the peripheral nervous system, which connects the brain and spinal cord to the body, is composed of two parts: the somatic system (voluntary control) and the autonomic system (involuntary control; see Sherwood, 2010). Stress and environmental influences often invoke the autonomic system to protect the individual from harm or danger, even before consciously registered by the individual. Within the autonomic system, two opposing systems help regulate its state, including the sympathetic nervous system which causes a switch to a *fight-or-flight* response and the parasympathetic nervous system which causes a switch to a *rest-and-digest* response. In either case, nerve impulse, which includes the transportation of sodium ions and the use of voltage-gated sodium ion channels, is used to signal if the individual should have a *fight-or-flight* or a *rest-and-digest* reaction (Sherwood, 2010). In the case of the association results for overall temperament at weaning in this study, associated markers may be identifying genetic regions that influence the degree to which the individual responds and regulates its autonomic system to the handling and stress of being in a working pen environment.

Of the 147 windows identified as being associated for WBS force when using $\pi = 0$ (Appendix Table A-1), 25 have been reported in previous studies (Table 9). Many of these studies reported QTL locations in cM units, however, if the range of the QTL spanned over regions identified in this study, the reference is listed. Enrichment analysis

Table 9. Chromosomal regions identified using BayesC with $\pi = 0$ analysis of Warner-Bratzler shear force previously reported in literature

Chromosome	Megabase region	Reference
2	4	Casas et al. (1998)
4	20	Casas et al. (2001)
5	40	Casas et al. (2001)
10	15	McClure et al. (2012)
	40	McClure et al. (2012)
	86	Alexander et al. (2007)
	91	Alexander et al. (2007)
	100	Alexander et al. (2007)
11	1	McClure et al. (2012)
12	50	McClure et al. (2012)
13	67	McClure et al. (2012)
	75	McClure et al. (2012)
15	82	McClure et al. (2012)
16	11	McClure et al. (2012)
20	58	Casas et al. (2003)
	68	Casas et al. (2003)
21	59	McClure et al. (2012)
23	49	McClure et al. (2012)
25	2	McClure et al. (2012)
	31	Gutiérrez-Gil et al. (2008)
	35	Gutiérrez-Gil et al. (2008)
26	29	Wu et al. (2012) and McClure et al. (2012)
29	34	McClure et al. (2012)
	41	Casas et al. (2003) and Gutiérrez-Gill et al. (2008)
	44	Casas et al. (2003), Gutiérrez-Gill et al. (2008), and McClure et al. (2012)

was conducted for WBS force using 671 genes (located in associated windows) and a background list of 7,462 genes recognized and not duplicated using *Homo sapiens* information. Terms relating to extracellular region, its disassembly, and activity surrounding serine peptidases were identified based on significant *P*-values (Table 8). Similar to overall temperament at weaning, correction for multiple testing resulted in

some of the terms no longer being significant, but terms were still investigated to understand the biological role they may be identifying for WBS force.

Serine peptidases along with matrixins (also called matrix metallopeptidases responsible for connective tissue catabolism) are present in the extracellular matrix of cells (as reviewed by Sentandreu et al., 2002). Although studies have placed serine peptidases in muscle cells and preliminary evidence suggests they influence meat tenderness, little is known about their true mechanisms and role. What is known is that serine peptidases are part of a group of enzymes called Cathepsins, which include cysteine, aspartic, and serine peptidases, and interest in their potential role in meat tenderization is growing (e.g., Zamaro et al., 2005; Kemp et al., 2010).

Using young Charolais bulls, Zamaro et al. (2005) investigated the potential predictors of meat toughness including the perimortem concentration of cysteine and serine peptidase inhibitors. Conclusions from this study found that serine peptidase inhibitor levels were positively correlated to and predictive of toughness in these bulls. When combined with 5 other variables, they accounted for 70% of the animal variability for toughness in the longissimus muscle at 6 d post-mortem. Serpins, which is an acronym for serine proteinase inhibitors, is one of the most important families of serine peptidase inhibitors, but studies have found that serpins are not necessarily specific to serine peptidases (reviewed by Sentandreu et al., 2002). Ontology results from the present study further indicate that activity involving serine peptidases and, potentially their inhibitors are of interest in tenderness measured through WBS force.

Objective 2: Comparison of Models on Impact of Breeding Value Prediction

In the case of WBS force, the fixed effects factors included in the final model were originally type of cross and date of shear. In the preliminary analysis, inclusion of type of cross factor produced estimated genetic variance using ASReml software close to 0 ($\hat{\sigma}_g^2$ = 0.000 and h^2 = 0.000). This is mostly likely due to type of cross resembling pedigree information and could be inhibiting estimation of variances and heritability. Because of this, type of cross was removed from the model for further estimation procedures for the animal model, but was included for Bayesian analyses to ensure that stratification within marker genotypes due to population structure did not cause false associations (e.g., Lander and Schork, 1994; Marchini et al., 2004, Janss et al., 2012). Date of shear was included in all analyses pertaining to WBS force. Fixed effects used in in the training phase to predict breeding values for overall temperament included sex, family nested within sire, birth year-season combinations, and pen nested within birth year-season combinations.

The estimate of heritability for WBS force using the animal model fitting only the additive genetic component was 0.055 ± 0.091. This value is low in comparison to values previously reported for *Bos taurus* cattle, which have ranged from 0.11 to 0.53 depending on breed or cross and number of animals available for the study (Shackelford et al., 1994; reviewed by Minick et al., 2004), but studies using *Bos indicus* cattle have previously reported similar heritabilities due to an apparently low additive genetic component for this trait in *Bos indicus* cattle (e.g., Riley et al., 2003; Smith et al., 2007).

The estimate of heritability for overall temperament at weaning was 0.464 ± 0.167 using the animal model fitting only the additive genetic component. This was higher than posterior means of genomic heritability estimated through Bayesian methods (Table 10). Reported estimates of heritability for temperament measured through flight speed score have been reported as high as 0.40 (Burrow, 2001).

Estimated breeding values (EBV) from the animal model for WBS force and overall temperament at weaning were compared to genomic estimated breeding values (GEBV) from 3 Bayesian methods. BayesC method using $\hat{\pi}$ had the highest accuracy for both traits (Table 10). The simple linear regression coefficient ($\hat{\beta}_{y,x}$) of GEBV on phenotype was lowest for BayesB and the animal model for WBS force and overall temperament at weaning, respectively. Although the regression coefficient for the animal model was the lowest for overall temperament at weaning, indicating less bias, it had the lowest accuracy for either trait (Table 10). All analyses had significant Spearman rank correlation coefficients (Table 11) for either trait indicating that the amount of re-ranking may not be large. The animal model in comparison to Bayesian methods had the lowest Spearman rank correlations, indicating a higher number of individuals with breeding values changing rank. Quartile boundaries for each analysis of each trait are reported in Table 12.

The number of individuals whose breeding values changed more than 2 quartiles made up no more than 2% or 4% for WBS force or overall temperament at weaning, respectively (Table 13). In this case, however, only half of the population could change 3 quartiles between any two analyses, which would mean that of the 50% that could

Table 10. Genetic parameter estimates ($\hat{h}^2$, σ_g^2, $\sigma_{\hat{g}}^2$, and σ_{EBV}^2), prediction accuracies ($\hat{\rho}_{g\hat{g}}$), and linear regression coefficients ($\hat{\beta}_{y,x}$) for Warner-Bratzler shear force and overall temperament at weaning for 4 models[1]

	BayesC (π = 0)	BayesC with $\hat{\pi}$	BayesB with $\hat{\pi}$	Animal Model
Warner-Bratzler shear force				
$\hat{h}^2$	0.155	0.136	0.234	0.055
σ_g^2	0.064	0.056	0.097	0.023
$\sigma_{\hat{g}}^2$	0.054	0.047	0.084	0.020
σ_{EBV}^2	0.011	0.009	0.023	0.002
$\hat{\rho}_{g\hat{g}}$	0.787	0.795	0.775	0.717
$\hat{\beta}_{y,x}$	4.157	4.610	3.030	9.326
Overall temperament at weaning				
$\hat{h}^2$	0.210	0.206	0.227	0.464
σ_g^2	0.893	0.876	0.965	1.972
$\sigma_{\hat{g}}^2$	0.740	0.721	0.802	1.869
σ_{EBV}^2	0.182	0.193	0.232	0.620
$\hat{\rho}_{g\hat{g}}$	0.671	0.702	0.698	0.622
$\hat{\beta}_{y,x}$	2.996	2.885	2.646	1.924

[1]$\hat{h}^2$ is the estimate of heritability, where $\hat{h}^2$ for the Bayesian models is the posterior mean of the genomic heritability, σ_g^2 is the genetic variance calculated as $\hat{h}^2 * \sigma_P^2$, where $\sigma_P^2 = 0.413 \; or \; 4.253$ for Warner-Bratzler shear force or overall temperament at weaning, respectively, and is the estimate of phenotypic variance from the data. σ_{EBV}^2 is the variance of the estimated breeding values. Models used were Bayesian or traditional mixed model procedure, where $\hat{\pi}$ is 0.995 or 0.997 for Warner-Bratzler shear force or overall temperament at weaning, respectively.

Table 11. Spearman rank correlations between analyses for overall temperament at weaning and Warner-Bratzler shear force[1]

	BayesC (π = 0)	BayesC with $\hat{\pi}$	BayesB with $\hat{\pi}$	Animal model
BayesC (π = 0)		0.964	0.942	0.704
BayesC with $\hat{\pi}$	0.996		0.996	0.700
BayesB with $\hat{\pi}$	0.984	0.991		0.703
Animal Model	0.726	0.727	0.733	

[1]Analyses listed are the Bayesian or traditional mixed model procedure used to estimate breeding values, where $\hat{\pi}$ is 0.995 or 0.997 for Warner-Bratzler shear force (bottom diagonal) or overall temperament at weaning (top diagonal), respectively.

Table 12. Quartile boundaries for ranked estimated breeding values

Trait	Model	Minimum	Lower quartile (25th)	Median value (50th)	Upper quartile (75th)	Maximum
Overall temperament at weaning						
	BayesC with $\pi = 0$	-1.208	-0.287	-0.011	0.306	1.328
	BayesC with $\hat{\pi} = 0.997$	-1.249	-0.290	-0.007	0.307	1.355
	BayesB with $\hat{\pi} = 0.997$	-1.363	-0.319	-0.009	0.335	1.464
	Animal Model	-1.991	-0.588	0.119	0.658	2.116
Warner-Bratzler shear force						
	BayesC with $\pi = 0$	-0.263	-0.067	-0.003	0.056	0.378
	BayesC with $\hat{\pi} = 0.995$	-0.231	-0.059	-0.003	0.050	0.337
	BayesB with $\hat{\pi} = 0.995$	-0.387	-0.104	-0.010	0.078	0.573
	Animal Model	-0.065	0.013	0.032	0.059	0.154

change, less than 3% and 8% of those animals did change 3 quartiles for WBS force and overall temperament at weaning, respectively. There was a tendency to see more animals have their breeding values change quartiles for overall temperament at weaning than WBS force, most likely due to the increased sample size. Between Bayesian analyses more than 70% of the animal's breeding values did not change at least a quartile for either trait, but less than 60% when comparing a Bayesian method to the animal model (Table 13). Overall, all Bayesian methods performed similarly, but BayesC using $\hat{\pi}$ had numerically higher accuracies and reliabilities than BayesB. For the remaining objectives, BayesB methods will be employed as it will allow comparisons to current literature and provided the lowest $\hat{\beta}_{y,x}$ for both traits.

Table 13. Comparison of the number of individuals with estimated breeding values for Warner-Bratzler shear force and overall temperament at weaning that changed *n* quartiles between any two analyses

	Number of individuals that changed *n* quartiles[1]			
Trait Models compared	0	1	2	3
Warner-Bratzler shear force				
BayesC ($\pi = 0$) vs. BayesC ($\hat{\pi} = 0.995$)	359 (92.76%)	28 (7.24%)	0 (0%)	0 (0%)
BayesC ($\pi = 0$) vs. BayesB ($\hat{\pi} = 0.995$)	341 (88.11%)	46 (11.89%)	0 (0%)	0 (0%)
BayesC ($\pi = 0$) vs. Animal Model	203 (52.45%)	140 (36.18%)	40 (10.34%)	4 (1.03%)
BayesC ($\hat{\pi} = 0.995$) vs. BayesB ($\hat{\pi} = 0.995$)	351 (90.70%)	36 (9.30%)	0 (0%)	0 (0%)
BayesC ($\hat{\pi} = 0.995$) vs. Animal Model	205 (52.97%)	141 (36.43%)	36 (9.30%)	5 (1.29%)
BayesB ($\hat{\pi} = 0.995$) vs. Animal Model	203 (52.45%)	144 (37.21%)	38 (9.82%)	2 (0.52%)
Overall temperament at weaning				
BayesC ($\pi = 0$) vs. BayesC ($\hat{\pi} = 0.997$)	594 (77.24%)	174 (22.63%)	1 (0.13%)	0 (0%)
BayesC ($\pi = 0$) vs. BayesB ($\hat{\pi} = 0.997$)	547 (71.13%)	218 (28.35%)	4 (0.52%)	0 (0%)
BayesC ($\pi = 0$) vs. Animal Model	320 (41.61%)	312 (40.57%)	109 (14.17%)	28 (3.64%)
BayesC ($\hat{\pi} = 0.997$) vs. BayesB ($\hat{\pi} = 0.997$)	713 (92.72%)	56 (7.28%)	0 (0%)	0 (0%)
BayesC ($\hat{\pi} = 0.997$) vs. Animal Model	317 (41.22%)	327 (42.52%)	98 (12.74%)	27 (3.51%)
BayesB ($\hat{\pi} = 0.997$) vs. Animal Model	311 (40.44%)	342 (44.47%)	90 (11.70%)	26 (3.38%)

[1]The number of quartiles changed was calculated by first assigning an animal's quartile for any given analysis based on Table 11, then finding the difference of each animal's quartile between the two analyses compared. Percentage was calculated by dividing the number of individuals within that category by the total number of animals (n = 389 or 769 for Warner-Bratzler shear force or overall temperament at weaning, respectively).

Objective 3: Inclusion and Impact of Breed-of-Origin Genotype on Breeding Value Prediction

Effects of marker alleles based on their ability to segregate with the causative allele could be affected by the breed from which it originated, indicating a breed- or even parent-of-origin effect. Markers with breed-of-origin genotypes mapped on autosomes for the F_2 ET offspring of Cycle 1 were used in this study (n = 34,449).

Unassigned Breed-of-Origin. Within the 469 F_2 ET animals that had breed-of-origin genotypes assigned (i.e., 0, 1, or 2 based on the number of N alleles), there was on average 436 regions that had either one or more alleles unassigned (i.e., not assigned as A or N). The majority (~ 84%) of those regions had markers flanking the region with the same genotype, indicating that the missing region could be inferred to have the same origin as the flanking markers (e.g., a region of markers that genotyped 1 1 - - - - - 1 1 could be inferred to be 1 1 1 1 1 1 1 1 1, where "-" is a marker missing a breed-of-origin genotype and 1 is a marker genotyped as a heterozygote). The remaining regions (~16%) had flanking markers that were not genotyped as the same origin and were therefore replaced with the family's average genotype of that particular marker (e.g., a region of markers that genotyped 1 1 - - - - - 2 2 could not have the missing genotypes inferred in the same way and each unassigned marker was given the family's average breed-of-origin genotype for that marker).

There were some regions within each family that breed-of-origin was not assigned for one or both alleles and resulted in the family average genotype not having an assigned genotype. Among these regions, those with flanking markers assigned the same breed-of-origin genotypes (~ 89% of the regions, see Table 14) were inferred to have the same genotype (i.e., the same instance within animals) and the missing data was replaced with the breed-of-origin genotype of the flanking markers. In some cases (~ 11% of these family regions), the flanking markers did not have the same breed-of-origin genotype and each unassigned marker for the family's average genotype was replaced with a 1, the value of a heterozygote.

Table 14. Summary of marker regions missing breed-of-origin assignment on a per family basis

Family	Number of offspring	Average breed-of-origin genotype across available loci[1]	Number of regions per family that the ends were:[2]		Total number of regions missing genotypes	Average number of markers in a region	Range of missing genotypes in a region
			Same	Different			
70	33	1.2	374	29	403	8.7	1 - 34
71	63	1.2	266	20	286	8. 3	1 - 29
72	45	1.2	237	38	275	8.9	1 - 46
73	7	1.2	250	44	294	9.2	1 - 38
74	8	1.5	537	109	646	11.3	1 - 65
75	36	1.2	214	33	247	7.9	1 - 23
76	7	1.5	579	69	648	11.4	1 - 87
77	41	1.2	340	28	368	9.1	1 - 52
80	66	1.1	199	18	217	8.5	1 - 34
81	56	1.1	86	23	109	7.2	1 - 18
82	15	1.7	692	82	774	12.7	1 - 148
83	35	1.4	413	39	452	10.0	1 - 41
84	28	1.5	557	39	596	10.9	1 - 87
Average across families	33.8	1.311	364.9	43.9	408.8	9.5	1 - 54

[1]Genotype is based on the number of Nellore alleles, where it could be 0, 1, or 2, and a 1 would indicate a heterozygote, that is, one allele each of Angus and Nellore origin.

[2]"Same" refers to flanking markers of the region with missing genotypes being the same (e.g., 1 1 - - - - - 1 1) and "Different" refers to the flanking markers of the region with missing genotypes being different (e.g., 1 1 - - - - - 2 2).

On average, across markers within a family, the breed-of-origin genotype across all 34,449 markers that were assigned was approximately the value of a heterozygote (Table 14). This was true for most families except 74, 76, 82, and 84. These first three families had among the fewest number of progeny and were most likely influenced by sampling error. Replacing missing genotypes of a family's average genotype with the value of a heterozygote appeared to be the most appropriate approach. There is the potential that this could influence prediction results, but these assumptions would permit the use of all the markers. In the event that the region was commonly unassigned across families, all animals would have a breed-of-origin genotype of 1 (i.e., a heterozygote) and association of that marker's breed-of-origin genotype is not likely to be found as it would not be segregating within the population, thereby not adding information to the prediction equations.

Breeding Value Prediction and Comparison. Of the original 441 F_2 ET animals with overall temperament at weaning scores used in Objective 2, 440 had breed-of-origin genotypes. Breeding values, utilizing breed-of-origin genotypes with or without the original nucleotide-based genotypes, were estimated using BayesB with $\hat{\pi}$ = 0.997, as concluded in Objective 2. Family nested within sire effect was not included in these two analyses to avoid using effects already accounted for in the breed-of-origin genotypes (based on assignment techniques). These same animals were used to run an additional analysis using only the nucleotide-based genotypes for easy comparison to the analysis that included breed-of-origin genotypes, but family nested within sire was included in

the model along with sex, birth year-season combinations and pen nested within birth year-season combinations.

Ideally, the combined set of markers would require $\hat{\pi} = \left(\frac{1}{2}\right) * 0.997$ to ensure that the same number of markers were included in the model compared to using only nucleotide-based or breed-of-origin-based genotypes. For this study, however, $\hat{\pi}$ was set to 0.997 in all cases, but will be addressed in the future. Although the nucleotide-based and breed-of-origin-based analyses resulted in similar genomic heritability estimates, the accuracy using only breed-of-origin genotypes instead of nucleotide-based genotypes were much higher and the regression parameter was closer to 1 (Table 15). When combining both genotype sets, there was an improvement in accuracy over using breed-of-origin-based genotypes, and further reduction of bias based on the regression parameter being closer to 1 (Table 15).

Table 15. Genetic parameter estimates ($\hat{h}^2, \sigma_g^2, \sigma_{\hat{g}}^2,$ and σ_{EBV}^2), prediction accuracies ($\hat{\rho}_{g\hat{g}}$), and simple linear regression coefficients ($\hat{\beta}_{y,x}$) for overall temperament at weaning using BayesB ($\hat{\pi}$ = 0.997) procedures with or without breed-of-origin genotypes[1]

	Genotypes used		
	Nucleotide-based	Breed-of-origin based	Combined set
$\hat{h}^2$	0.246	0.230	0.325
σ_g^2	1.053	0.981	1.387
$\sigma_{\hat{g}}^2$	0.803	0.842	1.160
σ_{EBV}^2	0.170	0.397	0.728
$\hat{\rho}_{g\hat{g}}$	0.531	0.816	0.876
$\hat{\beta}_{x,y}$	2.880	1.868	1.525

[1] $\hat{h}^2$ is the posterior mean of the genomic heritability, σ_g^2 is the genetic variance calculated as $\hat{h}^2 * \sigma_P^2$, where $\sigma_P^2 = 4.271$ for overall temperament at weaning, and is the estimate of phenotypic variance from the data. σ_{EBV}^2 is the variance of the estimated breeding values.

All instances resulted in positive and significant Spearman Rank correlations between the 3 analyses. Correlations were lowest (0.461) when comparing GEBV rank of nucleotide-based and the combined set and highest (0.825) when comparing GEBV rank of breed-of-origin genotypes only and the combined set. The percentage of breeding values that changed more than 2 quartiles was less than 5% in all cases (or less than 9% when considering the population that could change 3 quartiles), but only the comparison of breed-of-origin-based to the combined set resulted in over 50% of the GEBV not changing at least 1 quartile (Table 16, Table 17).

Ibáñez-Escriche et al. (2009) conducted a simulation study to look at the effects of modeling across-breed SNP genotypes (i.e., modeling it as a regular nucleotide-based genotypes) versus breed-specific SNP alleles, in order to predict purebred populations based on crossbred training population, with the intention that these purebred animals would be used to produce crossbred progeny. In their study, they accounted for the breed and parent in which it originated using BayesB procedures. Ibáñez-Escriche et al. (2009) found that as the relatedness between the breeds decreased and the number of breeds in the crossbred individuals used for training (either 2-, 3- or 4-breed crosses) increased, the accuracies of EBV decreased when modeling across-breed SNP genotypes or breed-specific SNP alleles. The breed-specific SNP alleles model typically performed better (i.e., had higher accuracies) than the across-breed SNP genotypes models, especially when the breeds were more distantly related and had fewer markers included in the model. This indicates that as the density of markers increases in an analysis, the need for breed-specific alleles in the model decreases.

Table 16. Quartile boundaries for overall temperament at weaning using BayesB procedures with or without breed-of-origin genotypes included in the model

Genotypes used	Minimum	Lower quartile (25th)	Median value (50th)	Upper quartile (75th)	Maximum
Nucleotide-based	-1.047	-0.302	-0.002	0.282	1.171
Breed-of-origin-based	-1.938	-0.404	0.001	0.463	1.810
Combined set	-2.294	-0.603	0.050	0.626	2.577

Table 17. Comparison of the number of individuals that changed *n* quartiles between any two analyses for overall temperament when considering breed-of-origin genotypes

	Number individuals that changed *n* quartiles[1]			
Models Compared	0	1	2	3
Nucleotide-based vs. Breed-of-origin-based	204 (46.36%)	175 (39.77%)	58 (13.18%)	3 (0.68%)
Nucleotide-based vs. Combined set	168 (38.18%)	175 (39.77%)	78 (17.73%)	19 (4.32%)
Breed-of-origin-based vs. Combined set	258 (58.64%)	148 (33.64%)	34(7.73%)	0 (0%)

[1]The number of quartiles changed was calculated by first assigning an animal's quartile for any given analysis based on Table 12, then finding the difference of each animal's quartile between the two analyses compared. Percentage was calculated by dividing the number of individuals within that category by the total number of animals (n = 440).

Similar results were found in the present study, as higher accuracies of GEBV were found using breed-of-origin-based genotypes over nucleotide-based genotypes. The combination of both types of genotypes, however, was not addressed in Ibáñez-Escriche et al. (2009), but was found to improve accuracy and lower bias over using nucleotide-based or breed-of-origin-base genotypes alone in the present study. Although performance of purebred individuals based on using crossbred animals in the training was not addressed in this study like Ibáñez-Escriche et al. (2009), there is potential to use the results from this study to impact the models and specifications in the commercial population. The procedures found in the present study were simpler to implement as

parent-of-origin was not taken into account. The assumptions made about missing breed-of-origin genotypes could have an impact on these results and it is of interest to determine that impact by investigating additional techniques to refine the assignment of breed-of-origin genotypes, perhaps by increasing the marker density to aid in phasing procedures on chromosomes. The improvement in accuracy and the lower bias using the combined set of genotypes over using nucleotide-based genotypes alone indicates that this method may be beneficial for composite or crossbred populations.

Objective 4: Impact of Dataset Partitioning

Several studies have demonstrated the importance of the relationship and structure of the training population in relation to the prediction population (e.g., de Roos et al., 2009; Ibáñez-Escriche et al., 2009; Habier et al., 2010b; Kizilkaya et al., 2010; Saatchi et al., 2010; Toosi et al., 2010). Saatchi et al. (2010) studied, specifically, the effect of the training record size and relationship of training to prediction population, and concluded that as the number of records increased, the accuracy increased and generations of training animals closer to the prediction animals yielded higher accuracies. Saatchi et al. (2010) and Habier et al. (2007) concluded that differences in accuracy based on training and prediction populations could be due to 1) the strength of relationship between the 2 populations, 2) the amount of recombination that can occur over time, and 3) the reduction of LD over generations.

Overall temperament at weaning (n = 769) was used to investigate the impact of relationship between the training and prediction population using a cross-validation

approach based on the population's family structure with BayesB ($\hat{\pi}$ = 0.997) and compared to EBV generated using an animal model. Training and validation groups for the animal model were created based on whether phenotype records for a particular group of animals were included in the analysis or not. Fixed effects to conduct training procedures included sex, family nested within sire, birth year-season combinations, and pen nested within birth year-season combinations for overall temperament at weaning.

The average additive relationship between the training population and the validation population was 0.044 across all 13 analyses and ranged from 0.001 to 0.308 (Table 18). There were 2 families that shared more limited relationship to the others based on the pedigree available, both of which were in Cycle 2 and consisted of reciprocal F_2 crosses.

When comparing the average across the 13 cross-validation training sets to the training set with all animals included (Table 10), the average accuracy was lower (Table 19), which could be due to the smaller sample sizes within each training set. Prediction accuracy was much lower in the validation groups (0.135 on average), but the regression coefficient was, on average, closer to 1 than the training groups (Tables 19 and 20). Interestingly, the use of progeny or grand-progeny by 297J resulted in the highest accuracies for GEBV although the maximum additive genetic relationship shared with the training group was the lowest compared to the other groups (Table 18). This could be due to markers capturing more relatedness between the training and validation groups than pedigree calculations.

In the event that grand-progeny were taken from the training group and placed in the validation group with full or half-siblings, there was a reduction in accuracy, which was expected (e.g., Habier et al., 2010b). Overall, however, there was poor prediction of GEBV for animals in the validation populations, which could be due to the extent of linkage disequilibrium between markers and causative regions within this population, sample size, or both (Habier et al., 2007; Habier et al., 2010b). It would be of interest to determine the influence that including breed-of-origin genotypes may have on improving the prediction accuracy regardless of population size, which may be done in the future.

In comparison to EBV generated using an animal model, GEBV generated using BayesB procedures had higher accuracies in all cases regardless of comparing training or validation groups (Tables 19 and 20). In some cases, accuracy of EBV was negative indicating phenotypes and breeding values having opposite trends. In most cases, the regression parameter in the training population was closer to 1 for EBV than for GEBV, but this trend did not carry to prediction of breeding values in the validation groups. In conclusion, although prediction of breeding values using the Bayesian methods resulted in low accuracy values for the validation group, they were higher than using a traditional animal model and the bias associated with those predictions were, on average, closer to 1.

Table 18. Average and range of additive genetic relationships within training or validation populations and across populations

			Training population		Validation population		Across populations	
Analysis[1]	n_{Train}	$n_{Validation}$	Average	Range	Average	Range	Average	Range
Sire is 297J	595	174	0.124	0 - 0.581	0.291	0.250 - 0.500	0.019	0 - 0.135
Sire is 432H	583	186	0.118	0 - 0.581	0.269	0.250 - 0.531	0.035	0 - 0.508
Sire is 437J	577	192	0.108	0 - 0.581	0.304	0.250 - 0.533	0.028	0 - 0.291
Sire is 551G	612	157	0.106	0 - 0.578	0.330	0.250 - 0.516	0.023	0 - 0.190
Natural service offspring	441	328	0.125	0.004 - 0.533	0.084	0 - 0.500	0.069	0.004 - 0.250
Random Cycle 1 only	592	177	0.095	0 - 0.581	0.100	0.004 - 0.533	0.082	0 - 0.500
Sire or grandsire is 297J	588	181	0.125	0 - 0.581	0.283	0.127 - 0.508	0.017	0 - 0.135
Sire or grandsire is 432H	576	193	0.119	0 - 0.581	0.261	0.072 - 0.531	0.033	0 - 0.508
Sire or grandsire is 437J	566	203	0.109	0 - 0.581	0.292	0.082 - 0.578	0.026	0 - 0.291
Sire or grandsire is 551G	587	182	0.107	0 - 0.578	0.299	0.084 - 0.581	0.017	0 - 0.190
Non-F_2 offspring	461	308	0.118	0 - 0.533	0.090	0.004 - 0.500	0.071	0 - 0.250
Random all cycles	579	190	0.097	0 - 0.581	0.093	0 - 0.581	0.077	0 - 0.500
Cycles 2 and 3	699	70	0.098	0.004 - 0.533	0.098	0 - 0.454	0.071	0.004 - 0.250
Average	574	195	0.112		0.215		0.044	

[1]Analysis refers to the progeny selected to be in the validation population. The remaining animals were used in the training phase.

Table 19. Genetic parameter estimates ($\hat{h}^2, \sigma_g^2, \sigma_{\hat{g}}^2$, and σ_{EBV}^2), prediction accuracies ($\hat{\rho}_{g\hat{g}}$), and simple linear regression coefficients ($\hat{\beta}_{y,x}$) for 13 training groups used in cross-validation[1]

	BayesB ($\hat{\pi}$ = 0.997)						Animal model					
Analysis[2]	$\hat{h}_g^2$	σ_g^2	$\sigma_{\hat{g}}^2$	σ_{EBV}^2	$\hat{\rho}_{g\hat{g}}$	$\hat{\beta}_{y,x}$	$\hat{h}^2$	σ_g^2	$\sigma_{\hat{g}}^2$	σ_{EBV}^2	$\hat{\rho}_{g\hat{g}}$	$\hat{\beta}_{y,x}$
Sire is 297J	0.209	0.888	0.763	0.162	0.617	3.143	0.462	1.966	1.935	0.539	0.513	1.855
Sire is 432H	0.227	0.964	0.802	0.207	0.675	2.873	0.503	2.138	2.059	0.744	0.632	1.782
Sire is 437J	0.226	0.960	0.792	0.190	0.667	3.052	0.557	2.370	2.302	0.952	0.610	1.497
Sire is 551G	0.254	1.079	0.909	0.290	0.746	2.545	0.506	2.153	2.097	0.681	0.673	2.098
Natural service offspring	0.247	1.051	0.804	0.170	0.534	2.878	0.231	0.982	0.761	0.044	0.389	7.649
Random (cycle 1 only)	0.258	1.098	0.936	0.300	0.664	2.241	0.537	2.283	2.245	0.820	0.608	1.678
Sire or grandsire is 297J	0.208	0.885	0.765	0.160	0.618	3.170	0.492	2.094	2.102	0.674	0.524	1.631
Sire or grandsire is 432H	0.236	1.005	0.838	0.233	0.684	2.697	0.591	2.513	2.534	1.353	0.672	1.253
Sire or grandsire is 437J	0.224	0.952	0.770	0.173	0.602	2.987	0.579	2.462	2.395	0.689	0.508	1.790
Sire or grandsire is 551G	0.268	1.141	0.949	0.323	0.718	2.312	0.568	2.416	2.369	0.791	0.641	1.940
Non-F2 offspring	0.258	1.098	0.849	0.198	0.540	2.637	0.717	3.050	3.260	0.830	0.467	1.774
Random (all cycles)	0.217	0.922	0.791	0.176	0.664	3.226	0.523	2.224	2.240	0.686	0.580	1.886
Cycles 2 and 3	0.229	0.972	0.778	0.206	0.640	2.703	0.537	2.285	2.178	0.508	0.529	2.325
Average	0.235	1.001	0.827	0.214	0.644	2.805	0.523	2.226	2.190	0.716	0.565	2.243

[1]$\hat{h}_g^2$ is the posterior mean of the genomic heritability, $\hat{h}^2$ is the heritability estimate from the animal model, σ_g^2 is the genetic variance calculated as $\hat{h}^2 * \sigma_P^2$, where $\sigma_P^2 = 4.253$ for overall temperament at weaning, and is the estimate of phenotypic variance from the data. σ_{GEBV}^2 is the variance of the estimated breeding values.

[2]Analysis refers to the progeny selected to be in the validation population. The remaining animals were used in the training phase.

Table 20. Genetic parameter estimates ($\hat{h}^2, \sigma_g^2, \sigma_{\hat{g}}^2$, and σ_{EBV}^2), prediction accuracies ($\hat{\rho}_{g\hat{g}}$), and simple linear regression coefficients ($\hat{\beta}_{y,x}$) for 13 validation groups used in cross-validation[1]

	BayesB ($\hat{\pi}$ = 0.997)						Animal model					
Analysis[2]	$\hat{h}_g^2$	σ_g^2	$\sigma_{\hat{g}}^2$	σ_{EBV}^2	$\hat{\rho}_{g\hat{g}}$	$\hat{\beta}_{y,x}$	$\hat{h}^2$	σ_g^2	$\sigma_{\hat{g}}^2$	σ_{EBV}^2	$\hat{\rho}_{g\hat{g}}$	$\hat{\beta}_{y,x}$
Sire is 297J	0.209	0.888	0.667	0.062	0.246	3.039	0.462	1.966	1.935	0.087	0.108	2.438
Sire is 432H	0.227	0.964	0.706	0.092	0.081	0.730	0.503	2.138	2.059	0.100	0.000	-0.004
Sire is 437J	0.226	0.960	0.670	0.084	0.178	1.695	0.557	2.370	2.302	0.154	-0.003	-0.047
Sire is 551G	0.254	1.079	0.784	0.183	0.099	0.499	0.506	2.153	2.097	0.024	-0.014	-1.261
Natural service offspring	0.247	1.051	0.849	0.090	0.085	0.891	0.231	0.982	0.761	0.003	0.005	1.231
Random (cycle 1 only)	0.258	1.098	0.939	0.242	0.188	0.788	0.537	2.283	2.245	0.355	0.107	0.683
Sire or grandsire is 297J	0.208	0.885	0.674	0.063	0.231	2.833	0.492	2.094	2.102	0.106	0.104	2.062
Sire or grandsire is 432H	0.236	1.005	0.722	0.096	0.062	0.555	0.591	2.513	2.534	0.150	-0.008	-0.130
Sire or grandsire is 437J	0.224	0.952	0.687	0.089	0.164	1.486	0.579	2.462	2.395	0.158	-0.006	-0.094
Sire or grandsire is 551G	0.268	1.141	0.793	0.171	0.115	0.643	0.568	2.416	2.369	0.057	-0.015	-0.656
Non-F2 offspring	0.258	1.098	0.877	0.112	0.070	0.612	0.717	3.050	3.260	0.060	0.004	0.219
Random (all cycles)	0.217	0.922	0.794	0.121	0.143	1.009	0.523	2.224	2.240	0.240	0.076	0.706
Cycles 2 and 3	0.229	0.972	0.761	0.158	0.093	0.505	0.537	2.285	2.178	0.139	0.032	0.519
Average	0.235	1.001	0.763	0.120	0.135	1.176	0.523	2.226	2.190	0.126	0.030	0.436

[1] $\hat{h}_g^2$ is the posterior mean of the genomic heritability, $\hat{h}^2$ is the heritability estimate from the animal model, σ_g^2 is the genetic variance calculated as $\hat{h}^2 * \sigma_P^2$, where $\sigma_P^2 = 4.253$ for overall temperament at weaning, and is the estimate of phenotypic variance from the data. σ_{GEBV}^2 is the variance of the estimated breeding values.

[2] Analysis refers to the progeny selected to be in the validation population. The remaining animals were used in the training phase.

CHAPTER V

CONCLUSION

For both WBS force and overall temperament at weaning, genomic regions were identified as associated when $\pi = 0$, which assumes an infinitesimal type model, but not when using $\hat{\pi}$. Ontology searches identified sodium-gated and voltage-gated ion channels relating to nervous system response as possible genetic mechanisms for overall temperament at weaning, and the role of serine peptidases, serine peptidase inhibitors, or both as possible genetic mechanisms for WBS force. In comparing models to predict breeding values, BayesC using $\hat{\pi}$ performed more accurately for WBS force and overall temperament at weaning, although there were small differences between it and BayesB. Inclusion of breed-of-origin genotypes in the BayesB model ($\hat{\pi} = 0.997$) either by itself or with the nucleotide-based genotypes (i.e., combined set) for overall temperament improved the accuracy and reduced the regression coefficient compared to using only nucleotide-based genotypes. This method was simple to implement after breed-of-origin genotypes were assigned and potentially could improve prediction accuracy in crossbred or composite breeds within the commercial cattle industry. The assignment of breed-of-origin genotypes, however, was not a trivial process and further investigation into techniques to refine assignment of breed-of-origin within commercial populations is of interest to minimize the number of unassigned genotypes. Cross-validation techniques resulted in low prediction accuracies for this study when using Bayesian methods, most likely influenced by sample size, but did perform better in validation groups than EBV

generated using the animal model. It is of interest to determine if including breed-of-origin genotypes with these populations' nucleotide-based genotypes would improve accuracy regardless of training population size, which can be done in the future.

REFERENCES

Alexander, L. J., M. D. MacNeil, T. W. Geary, W. M. Snelling, D. C. Rule, and J. A. Scanga. 2007. Quantitative trait loci with additive effects on palatability and fatty acid composition of meat in a Wagyu-Limousin F_2 population. Animal Genetics 38: 506-513.

Amen, T. S. 2007. Feed efficiency, carcass and temperament traits in F_2 Nellore-Angus steers, Ph.D. Dissertation, Texas A&M University, College Station.

Behrends, S. M., R. K. Miller, F. M. Rouquette Jr, R. D. Randel, B. G. Warrington, T. D. A. Forbes, T. H. Welsh, H. Lippke, J. M. Behrends, G. E. Carstens and J. W. Holloway. 2009. Relationship of temperament, growth, carcass characteristics and tenderness in beef steers. Meat Science 83(3): 433-438.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B (Methodological) 1995: 289-300.

Boddicker, N., E. H. Waide, R. R. R. Rowland, J. K. Lunney, D. J. Garrick, J. M. Reecy, and J. C. M. Dekkers. 2012. Evidence for a major QTL associated with host response to Porcine Reproductive and Respiratory Syndrome Virus challenge. J. Anim. Sci. 90: 1733-1746.

Boldt, C. R. 2008. A study of cattle disposition: exploring QTL associated with temperament. University Undergraduate Research Thesis, Texas A&M University, College Station.

Bolormaa, S., L. R. Porto Neto, Y. D. Zhang, R. J. Bunch, B. E. Harrison, M. E. Goddard, and W. Barendse. 2011. A genome-wide association study of meat and carcass traits in Australian cattle. J. Anim. Sci. 89: 2297-2309.

Brandt, A. E. 1933. The analysis of variance in a "2 x s" table with disproportionate frequencies. J. Am. Stat. Assoc. 28: 164-173.

Bulmer, M. G. 1980. The mathematical theory of quantitative genetics. Claredon Press, Oxford.

Burrow, H. M. 2001. Variances and covariances between productive and adaptive traits and temperament in a composite breed of tropical beef cattle. Livestock Production Science 70: 213-233.

Burrow, H. M., G. W. Seifert, and N. J. Corbet. 1988. A new technique for measuring temperament in cattle. Proceedings of the Australian Society of Animal Production 17: 154-157.

Calus, M. P. L. 2010. Genomic breeding value prediction: methods and procedures. Animal 4: 157-164.

Calus, M. P. L., T. H. E. Meuwissen, J. Windig, C. Schrooten, A. Vereijken, and R. F. Veerkamp. 2009. Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. Gen. Sel. Evol. 41: 11.

Calus, M. P. L., and R. F. Veerkamp. 2007. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. J. Anim. Breed. Gen. 124: 362-368.

Calus, M. P. L., and R. F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. Gen. Sel. Evol. 43: 1-14.

Casas, E., J. W. Keele, S. D. Shackelford, M. Koohmaraie, T. S. Sonstegard, T. P. Smith, S. M. Kappes, and R. T. Stone. 1998. Association of the muscle hypertrophy locus with carcass traits in beef cattle. J. Anim. Sci. 76(2): 468-473.

Casas, E., S. D. Shackelford, J. W. Keele, M. Koohmaraie, T. P. Smith, and R. T. Stone. 2003. Detection of quantitative trait loci for growth and carcass composition in cattle. J. Anim. Sci. 81(12): 2976-2983.

Casas, E., R. T. Stone, J. W. Keele, S. D. Shackelford, S. M. Kappes, and M. Koohmaraie. 2001. A comprehensive search for quantitative trait loci affecting growth and carcass composition of cattle segregating alternative forms of the myostatin gene. J. Anim. Sci. 79(4): 854-860.

Daetwyler, H. D., B. Villaneuva, P. Bijma, and J. Woolliams. 2007. Inbreeding in genome-wide selection. J. Anim. Breed. Gen. 124: 369-379.

de Roos, A. P. W., B. J. Hayes, and M. E. Goddard. 2009. Reliability of genomic predictions across multiple populations. Genetics 183: 1545-1553.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Royal Stat. Soc. Series B (Methodological) 39: 1-38.

Eden, E., R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10:48.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. Fourth ed. Pearson Education Limited, Harlow, England.

Fernando, R. L., and D. J. Garrick. 2009. GenSel - User manual for a portfolio of genomic selection related analyses. Animal Breeding and Genetics, Iowa State University, IA, USA. http://taurus.ansci.iastate.edu/gensel.

Fernando, R. L., and D. J. Garrick. 2013. Bayesian methods applied to GWAS. In Genome-wide association studies and genomic predictions. C. Condo, J. H. J. Van der Werf, and B. J. Hayes, ed. Springer, Berlin. *In Press*.

Fernando, R. L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. Gen. Sel. Evol. 21: 467-477.

Fernando, R. L., D. Nettleton, B. R. Southey, J. C. M. Dekkers, M. F. Rothschild, M. Soller. 2004. Controlling the proportion of false positives in multiple dependent tests. Genetics 166: 611-619.

Fisher, R. A. 1922. On the mathematical foundations of theoretical statistics. Philosophical transactions of the Royal Society of London. Series A: containing papers of a mathematical or physical character. 222: 309-368.

Fisher, R. A. 1925. Statistical methods for research workers. Oliver & Boyd. Edinburgh.

Funkhouser, R. R. 2008. Evaluation of disposition scores in *Bos indicus*/*Bos taurus* cross calves at different stages of production, M.S. Thesis, Texas A&M University, College Station.

Garrick, D. J. 2010. The nature, scope, and impact of some whole-genome analyses in beef cattle. In: 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany.

Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi, and R. L. Fernando. 2009. Additive genetic variability and the Bayesian alphabet. Genetics 183: 347-363.

Gianola, D., G. de los Campos, O. Gonzalez-Recio, N. Long, H. Okut, G. J. M. Rosa, K. A. Weigel, and X. L. Wu. 2010. Statistical learning methods for genome-based analysis of quantitative traits. In: Proceedings of the 9th World Congress Applied to Livestock Production, 14: 1-6.

Gianola, D., B. Goffinet, and M. G. Bulmer. 1982. Sire evaluation with best linear unbiased predictors. Biometrics 38: 1085-1088.

Gianola, D., M. Perez-Enciso, and M. A. Toro. 2003. On marker-assisted prediction of genetic value: beyond the ridge. Genetics 163: 347-365.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thomson. 2009. ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. www.vsni.co.uk.

Goddard, M. E. 2009. Genomic selection: prediction accuracy and maximization of long term response. Genetica 136: 245-257.

Goffinet, B. 1983. Selection on selected records. Gen. Sel. Evol. 15: 91-98.

Grandin, T. 1993a. Behavioral agitation during handling of cattle is persistent over time. Applied Animal Behavior Science 36: 1-9.

Grandin, T. 1993b. Teaching principles of behavior and equipment design for handling livestock. J. Anim. Sci. 71: 1065-1070.

Gutiérrez-Gil, B., P. Wiener, G. R. Nute, D. Burton, J. L. Gill, J. D. Wood, and J. L. Williams. 2008. Detection of quantitative trait loci for meat quality traits in cattle. Animal genetics 39(1): 51-61.

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177: 2389-2397.

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2009. Genomic selection using low-density marker panels. Genetics 182: 343-353.

Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2010a. Extension of the Bayesian alphabet for genomic selection. In: 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany.

Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics 12: 186.

Habier, D., J. Tetens, F. R. Seefried, P. Lichtner, and G. Thaller. 2010b. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Gen. Sel. Evol. 42: 5.

Hall, N. L., D. S. Buchanan, V. L. Anderson, B. R. Ilse, K. R. Carlin, E. P. Berg. 2011. Working chute behavior of feedlot cattle can be an indication of cattle temperament and beef carcass composition and quality. Meat Science 89(1): 52-57.

Hartley, H. O., and J. N. K. Rao. 1967. Maximum-likelihood estimation for the mixed analysis of variance model. Biometrika 54:93-108.

Hayashi, T., and H. Iwata. 2010. EM algorithm for Bayesian estimation of genomic breeding values. BMC Genetics 11: 3.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: genomic selection in dairy cattle: progress and challenges. J. Dairy. Sci. 92: 433-443.

Hayes, B. J., and M. E. Goddard. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Gen. Sel. Evol. 33: 209-229.

Henderson, C. R. 1949. Estimates of changes in herd environment. J. Dairy Sci. 32: 706. (Abstr.).

Henderson, C. R. 1963. Selection index and expected genetic advance. In: W. D. Hanson and H. L. Robinson (eds.) Statistical Genetics and Plant breeding Symposium and Workshop. National Academy of Sciences – National Research Council, Washington, D. C.

Henderson, C. R. 1973. Sire evaluation and genetic trends. J. Anim. Sci. 1973: 10-41.

Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32: 69-83.

Henderson, C. R. 1990. Statistical methods in animal improvement: historical overview. Page 2-14 in Advances in Statistical Methods for Genetic Improvement of Livestock. D. Gianola and K. Hammond, ed. Spring-Verlag, Germany.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. M. von Krosigk. 1959. The estimation of environmental and genetic trends from records subject to culling. Biometrics 15: 192-218.

Hu, Z.-L., C. A. Park, X.-L. Wu, and J. M. Reecy. 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Research 41: D871-D879.

Ibáñez-Escriche, N., and A. Blasco. 2011. Modifying growth curve parameters by multitrait genomic selection. J. Anim. Sci. 89: 661-668.

Ibáñez-Escriche, N., R. L. Fernando, A. Toosi, and J. C. M. Dekkers. 2009. Genomic selection of purebreds for crossbred performance. Gen. Sel. Evol. 41: 12-22.

Janss, L., G. de los Campos, N. Sheehan, and D. Sorensen. 2012. Inferences from genomic models in stratified populations. Genetics 192: 693-704.

Kärkkäinen, H. P., and M. J. Sillanpää. 2012. Back to basics for Bayesian model building in genomic selection. Genetics 191: 969-987.

Kemp, C. M., P. L. Sensky, R. G. Bardsley, P. J. Buttery, and T. Parr. 2010. Tenderness – an enzymatic view. Meat Science 84: 248-256.

King, D. A., C. E. Schuehle Pfeiffer, R. D. Randel, T. H. Welsh Jr., R. A. Oliphint, B. E. Baird, K. O. Curley Jr., R. C. Vann, D. S. Hale, J. W. Savell. 2006. Influence of animal temperament and stress responsiveness on the carcass quality and beef tenderness of feedlot cattle. Meat Science 74(3): 546-556.

Kizilkaya, K., R. L. Fernando, and D. J. Garrick. 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. J. Anim. Sci. 88: 544-551.

Kolbehdari, D., L. R. Schaeffer, and J. A. B. Robinson. 2007. Estimation of genome-wide haplotype effects in half-sib designs. J. Anim. Breed. Gen. 124: 356-361.

Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124: 743-756.

Lander, E. S., and N. J. Schork. 1994. Genetic dissection of complex traits. Science 265: 2037-2048.

Lush, J. L. 1931. The number of daughters necessary to prove a sire. J. Dairy Sci. 14: 209-220.

Lush, J. L. 1933. The bull index problem in the light of modern genetics. J. Dairy Sci. 16: 501-522.

Marchini, J., L. R. Cardon, M. S. Phillips, and P. Donnelly. 2004. The effects of human population structure on large genetic association studies. Nature Genetics 36: 512-517.

McClure, M.C., H. R. Ramey, M. M. Rolf, S. D. McKay, J. E. Decker, R. H. Chapple, J. W. Kim, T. M. Taxis, R. L. Weaber, R. D. Schnabel, J. F. Taylor. 2012. Genome-wide association analysis for quantitative trait loci influencing Warner-Bratzler shear force in five taurine cattle breeds. Animal Genetics 43(6): 662-673.

Meuwissen, T. H. E., and M. E. Goddard. 1996. The use of marker haplotypes in animal breeding schemes. Gen. Sel. Evol. 28: 1-16.

Meuwissen, T. H. E., and M. E. Goddard. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185: 623-631.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829.

Meuwissen, T. H. E., T. R. Solberg, R. Shepherd, and J. A. Wooliams. 2009. A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. Genet. Sel. Evol. 41: 2.

Minick, J.A., M. E. Dikeman, E. J. Pollak, and D. E. Wilson. 2004. Heritability and correlation estimates of Warner-Bratzler shear force and carcass traits from Angus-, Charolais-, Hereford-, and Simmental-sired cattle. Canadian Journal of Animal Science 84: 599-609.

Moser, G., M. Khatkar, B. J. Hayes, and H. Raadsma. 2010. Accuracy of direct genomic values in Holstein bulls and cows using subset of SNP markers. Gen. Sel. Evol. 42: 37.

Nelder, J. A. 1968. The combination of information in generally balanced designs. J. R. Statist. Soc. B. 30: 303-311.

Nicholson, K. L. 2008. Meat quality and disposition of F2 Nellore x Angus cross cattle, Ph.D. Dissertation, Texas A&M University, College Station.

O'Conner, S. F., J. D. Tatum, D. M. Wulf, R. D. Green, and G. C. Smith. 1997. Genetic effects on beef tenderness in *Bos indicus* composite and *Bos taurus* cattle. J. Anim. Sci. 75: 1822-1830.

Onteru, S. K., B. Fan, Z-Q. Du, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2012. A whole-genome association study for pig reproductive traits. Animal Genetics 43: 18-26.

Onteru, S. K., B. Fan, M. T. Nikkilä, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. Whole-genome association analyses for lifetime reproductive traits in the pig. J. Anim. Sci. 89: 988-995.

Patterson, H. D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58: 545-554.

Pearson, K. 1903. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability of correlation of organs. Philosophical transactions of the Royal Society of London. Series A: containing papers of a mathematical or physical character. 200: 1-66.

Piyasatian, N., R. L. Fernando, and J. C. M. Dekkers. 2006. Genomic selection for composite line development using low density marker maps. In: 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, MG, Brasil.

Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32: 949-953.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Riley, D. G., C. C. Chase, A. C. Hammond, R. L. West, D. D. Johnson, T. A. Olson, and S. W. Coleman. 2003. Estimated genetic parameters for palatability traits of steaks from Brahman cattle. J. Anim. Sci. 81(1): 54-60.

Rius-Vilarrasa, E., R. F. Brøndum, I. Strandén, B. Guldbrandtsen, E. Strandberg, M. S. Lund, and W. F. Fikse. 2012. Influence of model specifications on the reliabilities of genomic prediction in a Swedish-Finnish red breed cattle population. J. Anim. Breed. Gen. 2012: 1-11.

Rolf, M. M., J. F. Taylor, R. D. Schnabel, S. D. McKay, M. C. McClure, S. L. Northcutt, M. Kerley, and R. Weaber. 2010. Impact of reduced marker set estimation of genomic relationship matrices on genomic selection for feed efficiency in Angus cattle. BMC Genetics 11: 24.

Saatchi, M., M. C. McClure, S. D. McKay, M. M. Rolf, J. Kim, J. E. Decker, R. M. Taxis, R. H. Chapple, H. R. Romey, S. L. Northcutt, S. Bauck, B. Woodward, J. C. M. Dekkers, R. L. Fernando, R. D. Schnabel, D. J. Garrick, and J. F. Taylor. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Gen. Sel. Evol. 43: 40.

Saatchi, M., S. R. Miraei-Ashtiani, A. Nejati Javaremi, M. Moradi-Shahrebabak, and H. Mehrabani-Yeghaneh. 2010. The impact of information quantity and strength of relationship between training set and validation set on accuracy of genomic estimated breeding values. African J. Biotech. 9: 438-442.

Saatchi, M., R. D. Schnabel, M. M. Rolf, J. F. Taylor, and D. J. Garrick. 2012. Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. Gen. Sel. Evol. 44: 1-10.

Saatchi, M., J. Ward, and D. J. Garrick. 2013. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. J. Anim. Sci., published ahead of print doi:10.2527/jas.2012-5593.

Satagopan, J. M., and B. S. Yandell. 1996. Estimating the number of quantitative trait loci via Bayesian model determination. Proc. Joint Statistical Meetings, Chicago, IL.

Satagopan, J. M., B. S. Yandell, M. A. Newton, and T. C. Osborn. 1996. A Bayesian approach to detect quantitative trait loci using Markov Chain Monte Carlo. Genetics 144: 805-816.

Sentandreu, M. A., G. Coulis, and A. Ouali. 2002. Role of muscle endopeptidases and their inhibitors in meat tenderness. Trends in Food Science & Technology 13: 400-421.

Shackelford, S. D., M. Koohmaraie, L. V. Cundiff, K. E. Gregory, G. A. Rohrer, J. W. Savell. 1994. Heritabilities and phenotypic and genetic correlations for bovine postrigor calpastatin activity, intramuscular fat content, Warner-Bratzler shear force, retail product yield, and growth rate. J. Anim. Sci. 72(4): 857-863.

Sheet, P. and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78: 629-644.

Shepherd, R., T. Meuwissen, and J. Woolliams. 2010. Genomic selection and complex trait prediction using a fast EM algorithm applied to genome-wide markers. BMC Bioinformatics 11: 529.

Sherwood, L. 2010. Human physiology: from cells to systems. 7th ed. Brooks/Cole-Cengage Learning, Belmont, CA.

Sillanpää, M. J., and E. Arjas. 1998. Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. Genetics 148: 1373-1388.

Smith, T., J. D. Domingue, J. C. Paschal, D. E. Franke, T. D. Bidner, and G. Whipple. 2007. Genetic parameters for growth and carcass traits of Brahman steers. J. Anim. Sci. 85(6): 1377-1384.

Solberg, T. R., A. K. Sonesson, J. Woolliams, and T. H. E. Meuwissen. 2006. Genomic selection using different marker types and density. In: 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, MG, Brasil.

Solberg, T. R., A. K. Sonesson, J. A. Woolliams, and T. H. E. Meuwissen. 2008. Genomic selection using different marker types and densities. J. Anim. Sci. 86: 2447-2454.

Solberg, T. R., A. K. Sonesson, J. Woolliams, J. Odegard, and T. H. E. Meuwissen. 2009. Persistence of accuracy of genome-wide breeding values over generations when including a polygenic effect. Gen. Sel. Evol. 41: 53.

Stephens, D. A., and R. D. Fisch. 1998. Bayesian analysis of quantitative trait locus data using reversible jump Markov Chain Monte Carlo. Biometrics 54: 1334-1347.

Sun, X., D. Habier, R. L. Fernando, D. J. Garrick, and J. C. M. Dekkers. 2011. Genomic breeding value prediction and QTL mapping of QTLMAS2010 data using Bayesian methods. BMC Proceedings 5: S13.

ter Braak, C. J. F., M. P. Boer, and M. C. A. M. Bink. 2005. Extending Xu's Bayesian model for estimating polygenic effects using markers of the entire genome. Genetics 170: 1435-1438.

Toosi, A., R. L. Fernando, and J. C. M. Dekkers. 2010. Genomic selection in admixed and crossbred populations. J. Anim. Sci. 88: 32-46.

VanRaden, P. M. 2007. Genomic measures of relationship and inbreeding. In: Proceedings from 2007 Interbull Meeting, Dublin, Ireland.

VanRaden, P. M., and P. G. Sullivan. 2010. International genomic evaluation methods for dairy cattle. Gen. Sel. Evol. 42:7.

Veerkamp, R. F., K. L. Verbyla, H. A. Mulder, and M. P. L. Calus. 2010. Simultaneous QTL detection and genomic breeding value estimation using high density SNP chips. BMC Proceedings 4(Suppl 1): S9.

Verbyla, K. L., P. J. Bowman, B. J. Hayes, and M. E. Goddard. 2010. Sensitivity of genomic selection to using different prior distributions. BMC Proceedings 4: S5.

Verbyla, K. L., B. J. Hayes, P. J. Bowman, and M. E. Goddard. 2009. Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. Gen. Res. 91: 307-311.

Villumsen, T. M., L. Janss, and M. S. Lund. 2009. The importance of haplotype length and heritability using genomic selection in dairy cattle. J. Anim. Breed. Gen. 126: 3-13.

Wegenhoft, M. A. 2005. Locating quantitative trait loci associated with disposition in cattle. University Undergraduate Research Thesis, Texas A&M University, College Station.

Wright, S. 1921a. Correlation and causation. J. Agr. Res. 20: 557-585.

Wright, S. 1921b. Systems of mating. Genetics 6: 111-178.

Wright, S. 1932. On the evaluation of dairy sires. J. Anim. Sci. 1932: 71-78.

Wright, S. 1934. The method of path coefficients. The Annuals of Mathematical Statistics 5: 161-215.

Wu, X. X., Z. P. Yang, X. K. Shi, J. Y. Li, D. J. Ji, Y. J. Mao, L. L. Chang, and H. J. Gao. 2012. Association of SCD1 and DGAT1 SNPs with the intramuscular fat traits in Chinese Simmental cattle and their distribution in eight Chinese cattle breeds. Molecular Biology Reports 39(2): 1065-1071.

Xu, S. 2003a. Estimating polygenic effects using markers of the entire genome. Genetics 163: 789-801.

Xu, S. 2003b. Theoretical basis of the Beavis effect. Genetics 165: 2259-2268.

Yang, W., and R. J. Tempelman. 2010. A Bayesian antedependence model to account for linkage disequilibrium in whole genome selection. Page 0908 in Proc. 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany.

Yang, W., and R. J. Tempelman. 2012. A Bayesian antedependence model for whole genome prediction. Genetics. 190: 1491-1501.

Yates, F. 1934. The analysis of multiple classifications with unequal numbers in the different classes. J. Am. Stat. Assoc. 29: 51-66.

Yi, N., and S. Banerjee. 2009. Hierarchical generalized linear models for multiple quantitative trait locus mapping. Genetics 181: 1101-1113.

Yi, N., V. George, and D. B. Allison. 2003. Stochastic search variable selection for identifying multiple quantitative trait loci. Genetics 164: 1129-1138.

Yi, N., and D. Shriner. 2008. Advances in Bayesian multiple quantitative trait loci mapping in experimental crosses. Heredity 100: 240-252.

Yi, N., and S. Xu. 2008. Bayesian LASSO for quantitative trait loci mapping. Genetics 179: 1045-1055.

Yu, K., J. Xu, D. C. Rao, and M. Province. 2005. Using tree-based recursive partitioning methods to group haplotypes for increased power in association studies. Annals of Human Genetics 69: 577-589.

Zamora, F., L. Aubry, T. Sayd, J. Lepetit, A. Lebert, M. A. Sentandreu, A. Ouali. 2005. Serine peptidase inhibitors, the best predictor of beef ageing amongst a large set of quantitative variables. Meat Science 71(4): 730-742.

Zhan, B., J. Fadita, B. Thomsen, J. Hedegaard, F. Panitz, and C. Bendixen. 2011. Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. BMC Genomics 12(1): 557.

APPENDIX A

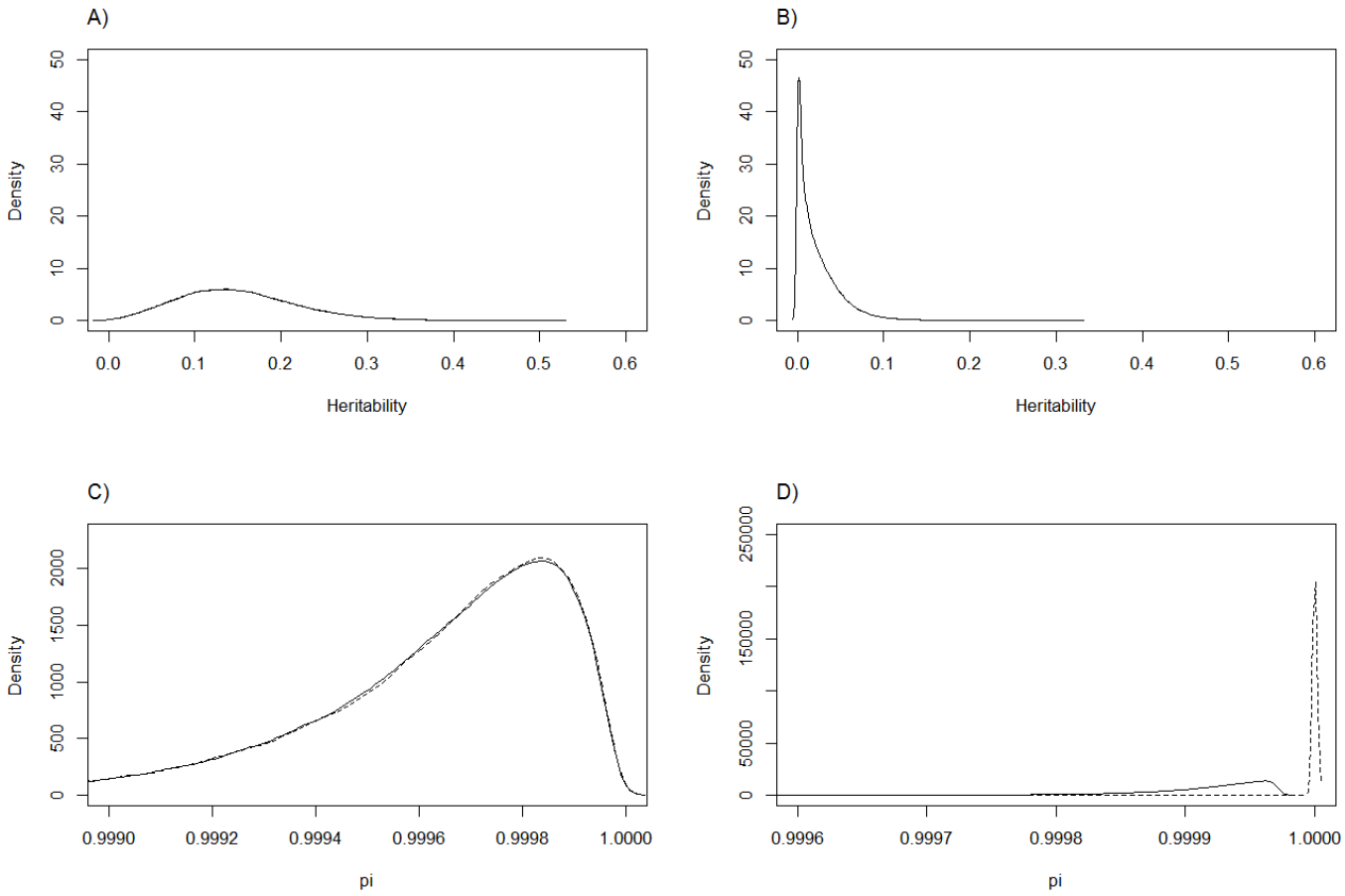

Figure A-1. Density plots of posterior estimates of MCMC samples using BayesCπ procedures with start values of $\pi = 0.1$ (solid line) or 0.9 (dashed line) for genomic heritability and π in overall temperament at weaning (A, C) and WBS force (B, D).

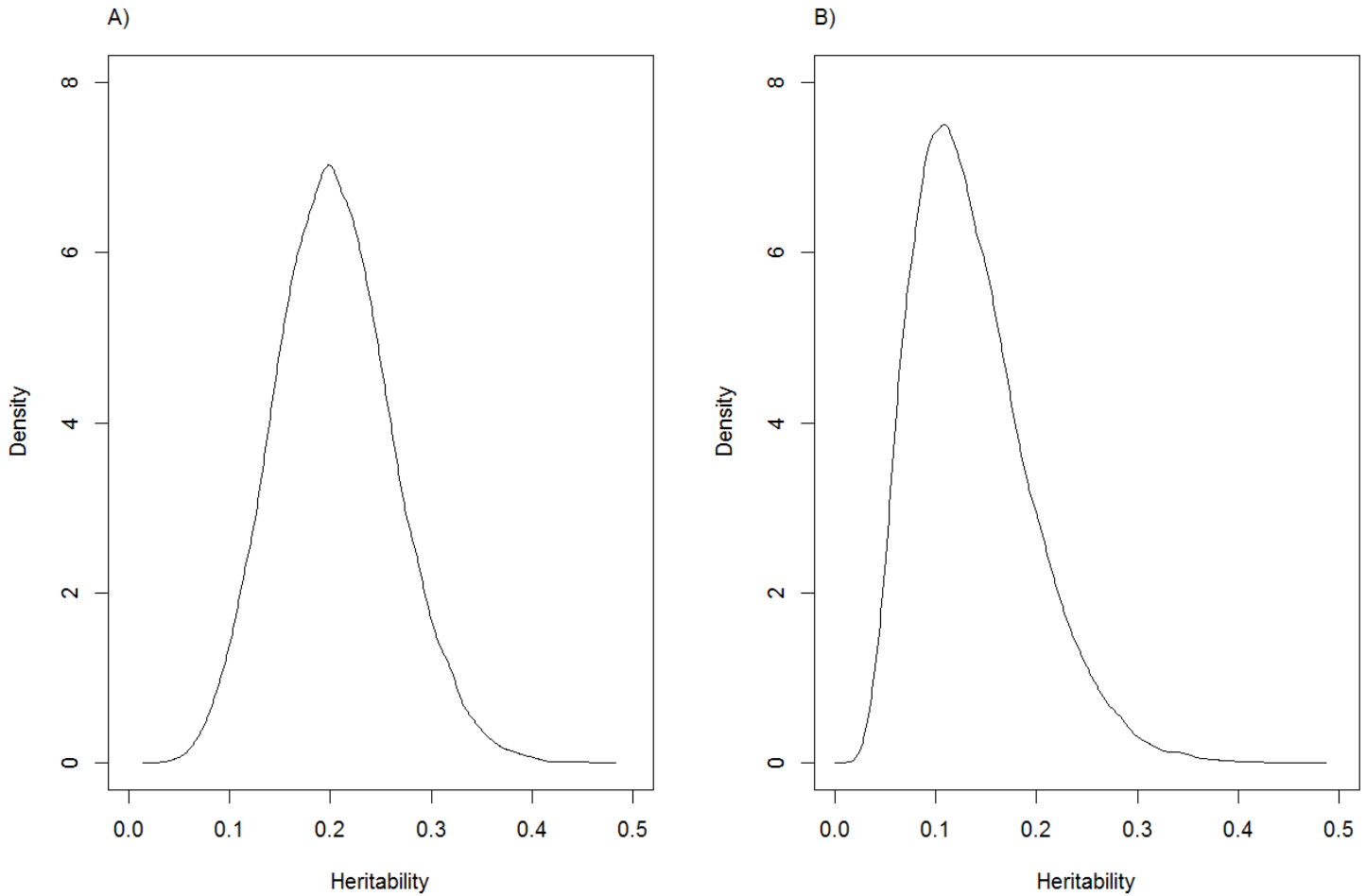

Figure A-2. Density plots of posterior estimates of MCMC samples using BayesC procedures for genomic heritability in overall temperament at weaning ($\hat{\pi}$ = 0.997; A) and WBS force ($\hat{\pi}$ = 0.995; B).

Association Results

Appendix Table A-1. Association results for overall temperament at weaning (OTW) and Warner-Bratzler Shear (WBS) force using Bayesian Inference with $\hat{\pi}$ when $PPA_w > 0.1$

					PPA_w				
					OTW		WBS Force		Identified when
Chromosome	Megabase Region	Number of SNP	Start Marker	End Marker	BayesB	BayesC	BayesB	BayesC	$\pi = 0$
1	40	22	BTB-01249999	ARS-BFGL-NGS-45335	-	-	0.1050	0.0986	Yes
	144	14	BTB-00067814	ARS-BFGL-NGS-17766	-	-	0.1028	0.0880	
	146	21	ARS-BFGL-NGS-105691	ARS-BFGL-NGS-113680	-	-	0.1018	0.0974	Yes
	147	22	Hapmap30204-BTA-124882	BTB-01761180	-	-	0.1054	0.0998	Yes
	152	19	Hapmap32785-BTA-124994	ARS-BFGL-NGS-77917	-	-	0.1018	0.0906	Yes
2	4	21	Hapmap43083-BTA-86781	ARS-BFGL-NGS-6152	0.2216	0.1922	-	-	Yes
	16	19	BTB-00080812	BTA-49719-no-rs	-	-	0.1008	0.0960	
	18	23	BTA-49599-no-rs	BTB-00082256	-	-	0.1334	0.1304	Yes
	21	22	Hapmap32159-BTA-46703	ARS-BFGL-NGS-32709	-	-	0.1238	0.1194	
	24	18	BTB-00086875	Hapmap40746-BTA-46710	-	-	0.1072	0.1068	
	28	24	Hapmap25643-BTA-47070	Hapmap43138-BTA-107007	-	-	0.1170	0.1196	Yes
	30	20	Hapmap25534-BTA-133814	ARS-BFGL-BAC-30866	-	-	0.1006	0.1048	Yes
	31	23	ARS-BFGL-BAC-33265	BTA-47279-no-rs	-	-	0.1274	0.1234	Yes
	55	28	Hapmap25908-BTA-160304	BTB-01160816	-	-	0.1276	0.1246	Yes
	78	25	ARS-BFGL-NGS-88183	ARS-BFGL-NGS-116036	-	-	0.1110	0.1192	Yes
	107	17	ARS-BFGL-NGS-102918	Hapmap23715-BTA-48904	-	-	0.1044	0.1026	
	112	19	BTA-34427-no-rs	BTA-49059-no-rs	-	-	0.1154	0.1086	
	114	19	ARS-BFGL-NGS-43428	Hapmap43218-BTA-26258	-	-	0.1100	0.1068	
	130	18	ARS-BFGL-NGS-102007	ARS-BFGL-NGS-43912	-	-	0.1046	0.0922	
	131	21	ARS-BFGL-NGS-54356	ARS-BFGL-NGS-77887	-	-	0.1432	0.1260	Yes
	133	24	ARS-BFGL-NGS-10870	ARS-BFGL-NGS-66860	-	-	0.1636	0.1428	Yes
	134	22	Hapmap29143-BTA-134433	ARS-BFGL-NGS-17681	-	-	0.1084	0.1124	Yes
	135	24	BTA-52274-no-rs	ARS-BFGL-NGS-63882	-	-	0.1374	0.1346	Yes
3	89	22	ARS-BFGL-NGS-10080	ARS-BFGL-NGS-44533	0.1070	0.0992	0.1090	0.1010	Yes
	103	16	ARS-BFGL-NGS-94424	Hapmap31146-BTA-151624	0.1010	0.1020	-	-	
	104	13	ARS-BFGL-NGS-35125	ARS-BFGL-NGS-1038	0.1262	0.1324	-	-	

Appendix Table A-1. Continued

Chromosome	Megabase Region	Number of SNP	Start Marker	End Marker	PPA_w OTW BayesB	PPA_w OTW BayesC	PPA_w WBS Force BayesB	PPA_w WBS Force BayesC	Identified when π = 0
3	108	18	ARS-BFGL-NGS-31442	ARS-BFGL-NGS-60487	0.1830	0.1450	-	-	
	109	19	ARS-BFGL-NGS-91446	Hapmap53978-rs29024239	0.2194	0.1862	-	-	
4	57	19	Hapmap55341-rs29010064	BTA-70744-no-rs	-	-	0.1076	0.0990	
	92	20	Hapmap42822-BTA-115746	Hapmap34400-BES9_Contig569_627	-	-	0.1108	0.1162	
	95	12	ARS-BFGL-NGS-75296	BTB-02090805	-	-	0.1296	0.0900	
	97	15	Hapmap25270-BTA-142450	ARS-BFGL-NGS-12738	0.1028	0.1002	-	-	
	103	19	BTA-25514-no-rs	ARS-BFGL-NGS-117579	-	-	0.1278	0.1000	Yes
	113	15	Hapmap57291-ss46526771	Hapmap22875-BTA-155031	-	-	0.1080	0.0930	
	119	22	BTB-01418405	ARS-BFGL-NGS-12243	-	-	0.1066	0.1076	Yes
5	8	22	BTB-01507733	BTB-00219231	-	-	0.1090	0.1074	Yes
	33	22	ARS-BFGL-NGS-110548	BTB-01495858	-	-	0.1200	0.1160	
	40	23	Hapmap55164-rs29016423	BTB-01494257	-	-	0.1122	0.1116	Yes
6	95	21	BTB-00270696	BTA-77380-no-rs	-	-	0.1028	0.0920	
7	2	22	ARS-BFGL-NGS-28272	ARS-BFGL-NGS-42465	-	-	0.1080	0.0956	Yes
	3	23	BTA-78493-no-rs	ARS-BFGL-NGS-10769	-	-	0.1164	0.1160	Yes
	5	20	ARS-BFGL-NGS-28126	ARS-BFGL-NGS-32423	-	-	0.1096	0.1002	
	65	19	ARS-BFGL-NGS-113181	ARS-BFGL-NGS-104938	-	-	0.1156	0.1068	Yes
8	24	21	BTB-00733178	BTB-01052583	-	-	0.1262	0.1162	
	34	19	BTB-01332284	BTB-01356443	-	-	0.1144	0.1134	
	37	21	Hapmap48192-BTA-114766	ARS-BFGL-NGS-68744	-	-	0.1292	0.1154	
	45	21	BTB-00344647	ARS-BFGL-NGS-113176	-	-	0.1076	0.0986	Yes
	50	19	Hapmap41647-BTA-81135	ARS-BFGL-NGS-111988	-	-	0.1142	0.1138	
	52	19	BTA-81238-no-rs	Hapmap41648-BTA-81180	-	-	0.1040	0.1036	
	55	23	Hapmap59214-rs29014649	BTA-119741-no-rs	-	-	0.1070	0.1112	Yes
	68	22	ARS-BFGL-NGS-88701	ARS-BFGL-NGS-2304	-	-	0.1046	0.1056	Yes
	107	18	Hapmap25843-BTA-146186	ARS-BFGL-NGS-20859	-	-	0.1148	0.1076	
10	13	23	ARS-BFGL-NGS-113038	ARS-BFGL-NGS-29028	-	-	0.1000	0.0992	Yes
	40	20	Hapmap60128-rs29023056	ARS-BFGL-BAC-12899	-	-	0.1034	0.0946	Yes
	44	24	BTB-01885735	Hapmap9514-BTA-67200	-	-	0.1014	0.1044	Yes
11	1	23	ARS-BFGL-NGS-21603	Hapmap54920-rs29026446	-	-	0.1038	0.1008	Yes

Appendix Table A-1. Continued

Chromosome	Megabase Region	Number of SNP	Start Marker	End Marker	PPA_w OTW BayesB	PPA_w OTW BayesC	PPA_w WBS Force BayesB	PPA_w WBS Force BayesC	Identified when π = 0
11	5	24	ARS-BFGL-NGS-69404	ARS-BFGL-NGS-34139	-	-	0.1014	0.1112	Yes
	6	20	ARS-BFGL-NGS-91251	ARS-BFGL-NGS-47869	-	-	0.1162	0.1052	Yes
	22	23	Hapmap23962-BTA-87546	BTB-00466422	-	-	0.1088	0.1070	Yes
	30	22	Hapmap38097-BTA-117206	ARS-BFGL-NGS-112032	-	-	0.1012	0.0978	Yes
	60	22	Hapmap52226-rs29018191	ARS-BFGL-NGS-110394	-	-	0.1124	0.1130	
12	1	10	Hapmap29441-BTA-160222	ARS-BFGL-NGS-104447	0.1130	0.1060	-	-	
	85	22	ARS-BFGL-NGS-64378	Hapmap31728-BTA-127926	-	-	0.1330	0.1218	
13	1	20	ARS-BFGL-NGS-71458	ARS-USMARC-Parent-EF026087-rs29011643	-	-	0.1084	0.1074	
	42	22	ARS-BFGL-NGS-64510	Hapmap54073-rs29012511	-	-	0.1098	0.1062	Yes
	46	25	BTB-00229660	BTB-01668668	-	-	0.1128	0.1156	Yes
14	1	10	Hapmap30381-BTC-005750	ARS-BFGL-NGS-71749	-	-	0.1456	0.1046	
	2	19	ARS-BFGL-NGS-107379	Hapmap24718-BTC-002945	-	-	0.1620	0.1294	
	3	15	Hapmap30375-BTC-003040	UA-IFASA-9288	-	-	0.1112	0.1024	
	4	23	Hapmap24777-BTC-064977	ARS-BFGL-NGS-109902	-	-	0.1308	0.1276	Yes
	5	24	ARS-BFGL-NGS-17644	UA-IFASA-9107	-	-	0.1274	0.1242	Yes
	6	22	Hapmap31968-BTC-056754	Hapmap33635-BTC-049051	-	-	0.1304	0.1186	Yes
	7	23	Hapmap27091-BTC-048823	ARS-BFGL-NGS-20781	-	-	0.1160	0.1104	Yes
	9	21	ARS-BFGL-NGS-110918	Hapmap24767-BTC-058058	-	-	0.1238	0.1246	
	10	27	Hapmap23517-BTC-058008	UA-IFASA-7013	-	-	0.1800	0.1664	Yes
	71	16	UA-IFASA-7753	Hapmap25804-BTA-129404	0.1530	0.1250	0.1188	0.0974	
	77	17	Hapmap34937-BES5_Contig538_1040	BTB-01243104	-	-	0.1148	0.0886	
15	44	21	BTA-36920-no-rs	BTB-01120359	-	-	0.1014	0.1054	Yes
	47	16	ARS-BFGL-NGS-19465	ARS-BFGL-NGS-2728	0.1116	0.1146	-	-	
	50	7	Hapmap33953-BES4_Contig513_1142	Hapmap61092-rs29026735	0.1006	0.0818	-	-	
	63	12	BTB-01746743	ARS-BFGL-NGS-41765	0.1110	0.1092	-	-	
	64	19	ARS-BFGL-NGS-82738	ARS-BFGL-NGS-84776	0.1248	0.1230	-	-	
	65	22	ARS-BFGL-NGS-43586	ARS-BFGL-NGS-65623	0.1002	0.1094	-	-	
	67	22	ARS-BFGL-NGS-107810	ARS-BFGL-NGS-106323	-	-	0.1038	0.1030	Yes

Appendix Table A-1. Continued

Chromosome	Megabase Region	Number of SNP	Start Marker	End Marker	PPA_w OTW BayesB	PPA_w OTW BayesC	PPA_w WBS Force BayesB	PPA_w WBS Force BayesC	Identified when $\pi = 0$
15	74	16	BTB-00619992	ARS-BFGL-NGS-119303	0.1170	0.1174	-	-	
	82	23	ARS-BFGL-NGS-1121	BTB-01665549	-	-	0.1070	0.1102	Yes
16	2	21	ARS-BFGL-NGS-53886	ARS-BFGL-NGS-10577	-	-	0.1276	0.1270	
	3	17	ARS-BFGL-NGS-28226	BTA-38126-no-rs	-	-	0.1546	0.1240	
	27	20	BTA-38252-no-rs	ARS-BFGL-NGS-62655	-	-	0.1046	0.1060	
	30	21	Hapmap53163-rs29012001	BTB-01274755	-	-	0.1174	0.1148	
	31	13	Hapmap29538-BTA-159421	ARS-BFGL-NGS-115293	-	-	0.1058	0.1020	
	33	20	BTB-00634325	Hapmap47277-BTA-38408	-	-	0.1492	0.1288	
	51	16	ARS-BFGL-NGS-111918	ARS-BFGL-NGS-41367	-	-	0.1268	0.1196	
	68	16	ARS-BFGL-NGS-101250	BTA-39797-no-rs	-	-	0.1068	0.0966	
	69	15	ARS-BFGL-NGS-53971	Hapmap59558-rs29017112	-	-	0.1278	0.0952	
	70	18	ARS-BFGL-NGS-116637	ARS-BFGL-NGS-29266	-	-	0.1012	0.0884	
17	10	19	Hapmap36789-SCAFFOLD190171_16889	BTB-02004836	-	-	0.1036	0.0988	Yes
	61	22	BTB-00681858	ARS-BFGL-NGS-7699	-	-	0.1070	0.1094	Yes
18	25	15	ARS-BFGL-NGS-24004	ARS-BFGL-NGS-109756	-	-	0.1498	0.0936	
19	13	16	Hapmap35067-BES1_Contig631_717	UA-IFASA-8072	-	-	0.1084	0.0804	
	14	19	ARS-BFGL-NGS-114182	ARS-BFGL-NGS-57566	-	-	0.1048	0.1112	
	23	15	ARS-BFGL-NGS-43245	Hapmap47955-BTA-44807	-	-	0.1060	0.0854	
	25	23	ARS-BFGL-NGS-84618	ARS-BFGL-NGS-93006	-	-	0.1226	0.1184	Yes
	49	22	ARS-BFGL-NGS-107345	UA-IFASA-5415	-	-	0.1378	0.1344	Yes
	54	21	ARS-BFGL-NGS-114192	ARS-BFGL-NGS-47707	-	-	0.1146	0.1056	Yes
	55	22	ARS-BFGL-NGS-116713	ARS-BFGL-NGS-16504	-	-	0.1064	0.0990	Yes
20	8	21	ARS-BFGL-NGS-35251	BTA-51494-no-rs	-	-	0.1012	0.1036	
	64	21	ARS-BFGL-NGS-22583	ARS-BFGL-NGS-111832	-	-	0.1020	0.1000	
21	12	22	ARS-BFGL-NGS-118231	Hapmap34873-BES4_Contig386_675	-	-	0.1008	0.1018	
	22	24	Hapmap41409-BTA-103368	ARS-BFGL-NGS-61826	-	-	0.1074	0.1114	
	45	11	ARS-BFGL-NGS-42945	Hapmap32804-BTA-135970	0.1304	0.1320	-	-	
	48	16	BTB-01480099	ARS-BFGL-BAC-29227	0.2971	0.2390	-	-	
22	3	17	ARS-BFGL-NGS-30499	ARS-BFGL-NGS-66359	-	-	0.1076	0.0992	
	5	20	Hapmap47040-BTA-54661	ARS-BFGL-NGS-33467	-	-	0.1032	0.1056	Yes

Appendix Table A-1. Continued

Chromosome	Megabase Region	Number of SNP	Start Marker	End Marker	PPA_w				Identified when $\pi = 0$
					OTW		WBS Force		
					BayesB	BayesC	BayesB	BayesC	
22	6	19	ARS-BFGL-NGS-34986	ARS-BFGL-NGS-66274	-	-	0.1082	0.1066	Yes
23	30	15	ARS-BFGL-NGS-26195	ARS-BFGL-NGS-37027	0.1208	0.1326	-	-	
	48	24	BTA-101164-no-rs	ARS-BFGL-NGS-20081	-	-	0.1010	0.1184	Yes
24	1	23	Hapmap47669-BTA-59022	ARS-BFGL-NGS-63597	-	-	0.1034	0.1034	Yes
	3	28	BTB-00876230	BTA-58748-no-rs	0.1090	0.1128	0.1266	0.1308	Yes
	13	12	ARS-BFGL-NGS-80841	BTA-34314-no-rs	0.1614	0.1580	-	-	
	21	18	Hapmap42845-BTA-121426	ARS-BFGL-NGS-10227	-	-	0.1058	0.0954	
	33	18	Hapmap56073-rs29020550	ARS-BFGL-NGS-54587	-	-	0.1146	0.0982	
25	11	22	ARS-BFGL-NGS-110934	ARS-BFGL-NGS-99730	-	-	0.1042	0.0968	Yes
	16	17	Hapmap42597-BTA-59441	ARS-BFGL-NGS-90323	0.1226	0.1036	-	-	
	21	15	Hapmap47058-BTA-59697	ARS-BFGL-NGS-103926	0.1144	0.1046	-	-	
	24	16	ARS-BFGL-NGS-17307	ARS-BFGL-NGS-97895	0.1726	0.1522	-	-	
	27	12	ARS-BFGL-NGS-111595	ARS-BFGL-NGS-20397	0.1408	0.1388	-	-	
	31	25	Hapmap30681-BTC-028325	Hapmap25329-BTA-159835	-	-	0.1112	0.1206	Yes
	34	22	ARS-BFGL-BAC-3741	ARS-BFGL-NGS-63889	-	-	0.1000	0.0990	
	35	23	ARS-BFGL-NGS-54279	ARS-BFGL-NGS-38544	-	-	0.1042	0.1084	Yes
26	36	15	BTB-01154553	BTB-01314304	-	-	0.1124	0.0970	
27	32	21	ARS-BFGL-NGS-72140	ARS-BFGL-NGS-106207	-	-	0.1022	0.0968	Yes
	41	20	Hapmap51908-BTA-63031	Hapmap27161-BTA-134658	-	-	0.1076	0.1060	
28	25	13	ARS-BFGL-NGS-111221	ARS-BFGL-NGS-13390	0.1000	0.0960	-	-	
	27	23	ARS-BFGL-NGS-102370	Hapmap47359-BTA-63927	-	-	0.1028	0.1026	Yes
29	3	14	BTB-01648891	UA-IFASA-9243	0.5067	0.3943	-	-	
	6	15	ARS-BFGL-NGS-43446	ARS-BFGL-NGS-15066	0.2026	0.2186	-	-	
	35	18	UA-IFASA-5914	ARS-BFGL-NGS-89027	-	-	0.1100	0.1040	Yes
	37	13	UA-IFASA-6120	ARS-BFGL-NGS-33015	-	-	0.1102	0.0970	
	41	22	ARS-BFGL-NGS-6059	ARS-BFGL-NGS-85356	-	-	0.1940	0.1546	Yes
	43	20	ARS-BFGL-NGS-118477	ARS-BFGL-NGS-16241	-	-	0.1782	0.1444	
	44	23	ARS-BFGL-NGS-21416	ARS-BFGL-NGS-34609	-	-	0.1476	0.1502	Yes
	46	20	ARS-BFGL-NGS-24800	ARS-BFGL-NGS-29253	-	-	0.1300	0.1160	
	48	19	ARS-BFGL-NGS-42102	ARS-BFGL-NGS-29493	-	-	0.1580	0.1294	
	49	12	ARS-BFGL-NGS-61470	Hapmap24687-BTA-152298	-	-	0.1326	0.1084	

Appendix Table A-2: Association results for overall temperament at weaning (OTW) and Warner-Bratzler shear (WBS) force using Bayesian Inference with $\pi = 0$ when $PPA_{w2} > 0.75$

					PPA_{w2}		
Chromosome	Megabase Region	Number of SNP	Start Marker	End Marker	OTW	WBS Force	Identified by Both
1	40	22	BTB-01249999	ARS-BFGL-NGS-45335	-	0.8186	
	41	20	BTB-01135743	BTB-00019496	-	0.8052	
	65	20	Hapmap31713-BTA-124020	BTB-00027583	-	0.7732	
	146	21	ARS-BFGL-NGS-105691	ARS-BFGL-NGS-113680	-	0.7924	
	147	22	Hapmap30204-BTA-124882	BTB-01761180	-	0.7870	
	152	19	Hapmap32785-BTA-124994	ARS-BFGL-NGS-77917	-	0.7698	
	155	20	Hapmap57477-rs29022938	ARS-BFGL-NGS-62440	-	0.8062	
2	4	21	Hapmap43083-BTA-86781	ARS-BFGL-NGS-6152	-	0.7796	
	18	23	BTA-49599-no-rs	BTB-00082256	-	0.8410	
	28	24	Hapmap25643-BTA-47070	Hapmap43138-BTA-107007	0.8146	0.8852	Yes
	30	20	Hapmap25534-BTA-133814	ARS-BFGL-BAC-30866	0.7518	0.8386	Yes
	31	23	ARS-BFGL-BAC-33265	BTA-47279-no-rs	-	0.8238	
	34	18	ARS-BFGL-NGS-65913	Hapmap45885-BTA-63051	-	0.7834	
	55	28	Hapmap25908-BTA-160304	BTB-01160816	0.8378	0.8956	Yes
	78	25	ARS-BFGL-NGS-88183	ARS-BFGL-NGS-116036	-	0.7784	
	131	21	ARS-BFGL-NGS-54356	ARS-BFGL-NGS-77887	-	0.7594	
	133	24	ARS-BFGL-NGS-10870	ARS-BFGL-NGS-66860	-	0.7848	
	134	22	Hapmap29143-BTA-134433	ARS-BFGL-NGS-17681	-	0.8048	
	135	24	BTA-52274-no-rs	ARS-BFGL-NGS-63882	0.8200	0.8748	Yes
3	85	21	BTA-68574-no-rs	BTB-01240613	-	0.7654	
	89	22	ARS-BFGL-NGS-10080	ARS-BFGL-NGS-44533	0.8024	0.8604	Yes
	98	22	ARS-BFGL-NGS-102345	ARS-BFGL-NGS-25489	-	0.7902	
4	20	21	ARS-BFGL-NGS-25363	Hapmap23330-BTA-159853	-	0.7548	
	36	18	Hapmap51645-BTA-85793	Hapmap28138-BTA-157033	-	0.8082	
	38	18	Hapmap46157-BTA-70132	UA-IFASA-6877	-	0.7520	
	66	18	ARS-BFGL-NGS-45911	ARS-BFGL-NGS-44674	-	0.7504	

Appendix Table A-2: Continued

Chromosome	Megabase Region	Number of SNP	Start Marker	End Marker	PPA_{w2} OTW	PPA_{w2} WBS Force	Identified by Both
	103	19	BTA-25514-no-rs	ARS-BFGL-NGS-117579	-	0.7922	
	112	16	Hapmap31474-BTA-152740	BTA-72569-no-rs	-	0.7624	
	117	19	BTB-01544170	ARS-BFGL-NGS-23165	-	0.7752	
	119	22	BTB-01418405	ARS-BFGL-NGS-12243	-	0.7746	
5	8	22	BTB-01507733	BTB-00219231	-	0.7668	
	40	23	Hapmap55164-rs29016423	BTB-01494257	0.7848	0.8610	Yes
6	7	16	Hapmap29564-BTA-143742	BTB-00242660	-	0.7600	
	13	21	BTB-01317228	BTA-77361-no-rs	-	0.7578	
	14	21	Hapmap24840-BTA-143762	Hapmap27184-BTA-149565	-	0.7754	
	41	22	BTB-00252896	BTA-75905-no-rs	-	0.7946	
	42	20	ARS-BFGL-NGS-90128	BTA-75926-no-rs	-	0.7974	
	43	21	Hapmap59861-rs29027897	BTB-01893222	-	0.8196	
	45	20	BTB-00254199	Hapmap48206-BTA-119876	-	0.7710	
	104	19	ARS-BFGL-NGS-114206	BTA-103976-no-rs	-	0.7530	
	106	20	ARS-BFGL-NGS-91817	BTB-01782372	-	0.7658	
	110	21	ARS-BFGL-NGS-100510	ARS-BFGL-NGS-12667	-	0.8230	
7	2	22	ARS-BFGL-NGS-28272	ARS-BFGL-NGS-42465	0.7774	0.8450	Yes
	3	23	BTA-78493-no-rs	ARS-BFGL-NGS-10769	-	0.8138	
	20	22	ARS-BFGL-NGS-83052	BTB-00296617	-	0.7534	
	65	19	ARS-BFGL-NGS-113181	ARS-BFGL-NGS-104938	-	0.7554	
	105	20	ARS-BFGL-NGS-23510	ARS-BFGL-NGS-104100	-	0.7856	
8	21	20	BTA-87492-no-rs	ARS-BFGL-NGS-84952	-	0.7602	
	45	21	BTB-00344647	ARS-BFGL-NGS-113176	-	0.7660	
	55	23	Hapmap59214-rs29014649	BTA-119741-no-rs	-	0.7790	
	68	22	ARS-BFGL-NGS-88701	ARS-BFGL-NGS-2304	0.8114	0.8696	Yes
	93	18	BTA-16616-no-rs	ARS-BFGL-NGS-18836	-	0.7672	
9	23	18	BTA-85354-no-rs	Hapmap46562-BTA-28711	-	0.7674	
	35	19	BTB-00389092	BTB-00518052	-	0.7940	
10	13	23	ARS-BFGL-NGS-113038	ARS-BFGL-NGS-29028	-	0.7876	
	15	18	ARS-BFGL-NGS-92657	Hapmap47826-BTA-107963	-	0.7896	

Appendix Table A-2: Continued

Chromosome	Megabase Region	Number of SNP	Start Marker	End Marker	PPA_{w2} OTW	PPA_{w2} WBS Force	Identified by Both
10	18	20	ARS-BFGL-NGS-94247	ARS-BFGL-NGS-32828	-	0.7794	
	40	20	Hapmap60128-rs29023056	ARS-BFGL-BAC-12899	-	0.7652	
	44	24	BTB-01885735	Hapmap9514-BTA-67200	0.7844	0.8442	Yes
	86	24	Hapmap40595-BTA-79304	ARS-BFGL-NGS-34516	-	0.7864	
	91	22	ARS-BFGL-NGS-53476	BTA-80379-no-rs	0.8132	0.8748	Yes
	100	22	Hapmap33658-BTA-125990	Hapmap54967-rs29014597	-	0.7794	
11	1	23	ARS-BFGL-NGS-21603	Hapmap54920-rs29026446	0.7748	0.8276	Yes
	3	18	ARS-BFGL-NGS-31804	ARS-BFGL-NGS-18450	-	0.7656	
	5	24	ARS-BFGL-NGS-69404	ARS-BFGL-NGS-34139	-	0.7954	
	6	20	ARS-BFGL-NGS-91251	ARS-BFGL-NGS-47869	-	0.7516	
	22	23	Hapmap23962-BTA-87546	BTB-00466422	0.7796	0.8312	Yes
	24	23	ARS-BFGL-NGS-35892	BTA-86248-no-rs	0.7558	0.8248	Yes
	30	22	Hapmap38097-BTA-117206	ARS-BFGL-NGS-112032	-	0.7858	
12	2	19	ARS-BFGL-NGS-74419	ARS-BFGL-BAC-15634	-	0.7658	
	16	22	ARS-BFGL-NGS-58789	ARS-BFGL-NGS-77196	-	0.7606	
	50	21	BTA-115896-no-rs	BTB-00705700	-	0.7680	
	53	19	ARS-BFGL-NGS-11432	BTB-00498136	-	0.7716	
13	42	22	ARS-BFGL-NGS-64510	Hapmap54073-rs29012511	-	0.7724	
	46	25	BTB-00229660	BTB-01668668	-	0.7508	
	56	21	ARS-BFGL-NGS-71271	Hapmap43556-BTA-33007	-	0.7898	
	67	21	ARS-BFGL-NGS-92308	Hapmap44949-BTA-33430	-	0.7672	
	71	18	ARS-BFGL-NGS-95635	ARS-BFGL-NGS-14434	-	0.7538	
	75	21	Hapmap31079-BTA-128662	ARS-BFGL-NGS-114018	-	0.8236	
	83	19	ARS-BFGL-NGS-19988	Hapmap40029-BTA-87430	-	0.8104	
14	4	23	Hapmap24777-BTC-064977	ARS-BFGL-NGS-109902	0.7774	0.8446	Yes
	5	24	ARS-BFGL-NGS-17644	UA-IFASA-9107	0.7576	0.8180	Yes
	6	22	Hapmap31968-BTC-056754	Hapmap33635-BTC-049051	-	0.8114	
	7	23	Hapmap27091-BTC-048823	ARS-BFGL-NGS-20781	0.8262	0.8818	Yes
	10	27	Hapmap23517-BTC-058008	UA-IFASA-7013	0.8118	0.8654	Yes
	27	24	BTB-01280026	Hapmap27563-BTC-073928	0.7616	0.8076	Yes
	56	21	Hapmap26539-BTC-012199	Hapmap30660-BTC-013711	-	0.7646	

Appendix Table A-2: Continued

Chromosome	Megabase Region	Number of SNP	Start Marker	End Marker	PPA_{w2} OTW	PPA_{w2} WBS Force	Identified by Both
15	23	22	ARS-BFGL-NGS-58474	Hapmap27194-BTA-153449	-	0.7890	
	26	21	BTA-96174-no-rs	BTB-00588098	-	0.7674	
	44	21	BTA-36920-no-rs	BTB-01120359	-	0.7570	
	67	22	ARS-BFGL-NGS-107810	ARS-BFGL-NGS-106323	-	0.7502	
	75	22	ARS-BFGL-NGS-119294	ARS-BFGL-NGS-111520	-	0.7604	
	82	23	ARS-BFGL-NGS-1121	BTB-01665549	0.8512	0.9050	Yes
16	11	19	BTA-39714-no-rs	ARS-BFGL-NGS-80278	-	0.7782	
	76	21	ARS-BFGL-NGS-119489	Hapmap47981-BTA-51932	-	0.8186	
17	9	20	BTA-42045-no-rs	Hapmap6261-BTA-42054	0.7506	0.8432	Yes
	10	19	Hapmap36789-SCAFFOLD190171_16889	BTB-02004836	-	0.7686	
	61	22	BTB-00681858	ARS-BFGL-NGS-7699	0.7556	0.8298	Yes
	62	23	ARS-BFGL-NGS-118025	ARS-BFGL-NGS-17397	-	0.7768	
	65	20	BTB-00675021	BTB-00683408	-	0.7752	
18	6	19	BTA-121111-no-rs	ARS-BFGL-NGS-93178	0.7882	0.8420	Yes
	23	19	Hapmap51823-BTA-31074	Hapmap22859-BTA-132194	0.7109	0.8024	
	26	22	ARS-BFGL-NGS-66258	UA-IFASA-7519	0.7784	0.8466	Yes
	64	19	BTB-01631910	ARS-BFGL-NGS-107405	-	0.7724	
19	25	23	ARS-BFGL-NGS-84618	ARS-BFGL-NGS-93006	0.7774	0.8392	Yes
	49	22	ARS-BFGL-NGS-107345	UA-IFASA-5415	0.7564	0.8014	Yes
	54	21	ARS-BFGL-NGS-114192	ARS-BFGL-NGS-47707	0.7744	0.8408	Yes
	55	22	ARS-BFGL-NGS-116713	ARS-BFGL-NGS-16504	-	0.8104	
	59	18	ARS-BFGL-NGS-12196	ARS-BFGL-NGS-22646	-	0.7640	
20	11	22	ARS-BFGL-NGS-71611	Hapmap52341-rs29025776	0.7792	0.8408	Yes
	39	20	ARS-BFGL-NGS-17676	ARS-BFGL-BAC-27243	-	0.7826	
	58	20	ARS-BFGL-NGS-90070	ARS-BFGL-NGS-71622	-	0.7734	
	68	20	ARS-BFGL-NGS-33801	ARS-BFGL-NGS-63132	-	0.8124	
21	19	23	ARS-BFGL-NGS-33883	Hapmap47651-BTA-51810	0.7882	0.8576	Yes
	35	17	ARS-BFGL-NGS-103555	BTB-00817049	-	0.7552	
	59	21	ARS-BFGL-NGS-2042	ARS-BFGL-NGS-79673	0.8312	0.8826	Yes

Appendix Table A-2: Continued

Chromosome	Megabase Region	Number of SNP	Start Marker	End Marker	PPA_{w2} OTW	PPA_{w2} WBS Force	Identified by Both
22	5	20	Hapmap47040-BTA-54661	ARS-BFGL-NGS-33467	-	0.8052	
	6	19	ARS-BFGL-NGS-34986	ARS-BFGL-NGS-66274	-	0.7574	
	52	21	ARS-BFGL-NGS-94797	ARS-BFGL-NGS-32152	-	0.7528	
23	14	20	ARS-BFGL-NGS-101393	ARS-BFGL-NGS-25325	0.7504	0.8366	Yes
	40	20	Hapmap27399-BTA-137408	ARS-BFGL-NGS-12710	-	0.8036	
	45	20	BTB-01877881	BTB-00867439	-	0.8226	
	48	24	BTA-101164-no-rs	ARS-BFGL-NGS-20081	0.8102	0.8704	Yes
	49	19	ARS-BFGL-NGS-61352	ARS-BFGL-NGS-137	-	0.7976	
24	1	23	Hapmap47669-BTA-59022	ARS-BFGL-NGS-63597	-	0.7590	
	3	28	BTB-00876230	BTA-58748-no-rs	0.8366	0.9006	Yes
	22	18	Hapmap38610-BTA-57584	ARS-BFGL-NGS-114549	-	0.7640	
	29	19	ARS-BFGL-NGS-35716	BTB-01721436	-	0.7836	
	32	19	ARS-BFGL-NGS-14224	ARS-BFGL-NGS-117214	-	0.7548	
25	2	17	Hapmap24734-BTC-016033	ARS-BFGL-NGS-10250	-	0.7620	
	5	20	Hapmap30085-BTC-001300	ARS-BFGL-NGS-26868	0.7524	0.8244	Yes
	11	22	ARS-BFGL-NGS-110934	ARS-BFGL-NGS-99730	0.8152	0.8730	Yes
	31	25	Hapmap30681-BTC-028325	Hapmap25329-BTA-159835	0.8582	0.9122	Yes
	35	23	ARS-BFGL-NGS-54279	ARS-BFGL-NGS-38544	-	0.7694	
26	10	20	Hapmap58185-rs29022254	ARS-BFGL-NGS-22284	-	0.7714	
	29	21	ARS-BFGL-NGS-60856	ARS-BFGL-NGS-116214	-	0.7638	
	45	22	ARS-BFGL-NGS-5906	ARS-BFGL-NGS-30060	0.7850	0.8442	Yes
	49	21	Hapmap27655-BTA-163523	ARS-BFGL-NGS-24474	-	0.7902	
27	14	21	ARS-BFGL-NGS-13588	Hapmap28258-BTA-139286	0.7634	0.8484	Yes
	32	21	ARS-BFGL-NGS-72140	ARS-BFGL-NGS-106207	-	0.8150	
28	27	23	ARS-BFGL-NGS-102370	Hapmap47359-BTA-63927	-	0.7568	
	34	21	ARS-BFGL-NGS-111706	ARS-BFGL-NGS-101535	-	0.7750	
29	34	18	BTA-66497-no-rs	ARS-BFGL-NGS-111197	-	0.7630	
	35	18	UA-IFASA-5914	ARS-BFGL-NGS-89027	-	0.7888	
	41	22	ARS-BFGL-NGS-6059	ARS-BFGL-NGS-85356	-	0.7992	
	44	23	ARS-BFGL-NGS-21416	ARS-BFGL-NGS-34609	-	0.8140	
X	113	12	Hapmap24068-BTA-112234	ARS-BFGL-NGS-66271	-	0.8100	

APPENDIX B

Objective 1 R Script: Plotting Association Windows

```
WBSF = read.table("PlacePathHere.txt", header=TRUE)
OvWnDp = read.table("PlacePathHere.txt", header=TRUE)

###Window QTL Plots###
#labeling reference points for plotting window QTL file
Npnts = 31 ##30 chromosomes, unknown, plus end point
N = dim(WBSF)[1]
Windows = matrix(0,npnts,2,dimnames=list(1:npnts,c("Min","Max")))

for(i in 1:npnts){
        Chr = subset(WBSF, chr = = i)
        Windows[i,'Min'] = min(Chr[,'Window'])
        Windows[i,'Max'] = max(Chr[,'Window'])
}
Ref = matrix(c(Windows[,'Min'],Windows[npnts,'Max']),npnts+1,1)
RefC = matrix(0,npnts,1)   #Center points to label with

for(i in 1:npnts){
        RefC[i,1] = round(((Ref[i+1,1]-Ref[i,1])/2)+Ref[i,1])
}

#par(mfrow = c(1,1), oma = c(0,0,0,2)) # Not necessary, but can place together
xrange = range(Ref[1:npnts,1])
yrange = c(0,1)

###Warner-Bratzler Shear Force###
###BC995 & BB995###
plot(xrange,yrange,yaxt="n",xaxt="n",xlab="Chromosome",ylab="PPAw",cex.lab=0.7,t
ype="n")

for(i in 1:(npnts-1)){
        Chromo = subset(WBSF, chr = = i)
        nWin = dim(Chromo)[1]
        for(j in 1:nWin){
                points(Chromo[j,1],Chromo[j,6],pch=16,col="gray70")
                points(Chromo[j,1],Chromo[j,7],pch=16,col="black")
        }
}
axis(2,las=1,cex.axis=0.7)
```

```
axis(1, at = Ref[1:31,1], labels =
c("","","","","","","","","","","","","","","","","","","","","","","","","","","","","","",""), tck
= 0.035, cex.axis = 0.8) #Creates blank labels at the end of chromosomes with tick
marks below axis

axis(1, at =Ref[1:31,1], labels =
c("","","","","","","","","","","","","","","","","","","","","","","","","","","","","","",""), tck
= -0.035, cex.axis = 0.8) #Creates blank labels at the end of chromosomes with tick
marks above axis

axis(1, at = RefC[1:npnts-1,1], labels =
c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,"X"),
tck = 0, cex.axis = 0.7) #Creates labels on the center of the chromosomes without tick
marks

legend("topright", legend = c("BayesB (0.995)", "BayesC (0.995)"), col =
c("gray70","black"), pch = c(16,16))

mtext("B)", side = 3, line = 1, adj = 0)


###Overall Temperament At Weaning###
###BB997 & BC997###
plot(xrange,yrange,yaxt="n",xaxt="n",xlab="Chromosome",ylab="PPAw",cex.lab=0.7,t
ype="n")

for(i in 1:(npnts-1)){
	Chromo = subset(OvWnDp, chr = = i)
	nWin = dim(Chromo)[1]
	for(j in 1:nWin){
		points(Chromo[j,1],Chromo[j,7],pch=16,col="gray70")
		points(Chromo[j,1],Chromo[j,6],pch=16,col="black")
	}
}
axis(2,las=1,cex.axis=0.7)

axis(1, at = Ref[1:31,1], labels =
c("","","","","","","","","","","","","","","","","","","","","","","","","","","","","","",""), tck
= 0.035, cex.axis = 0.8) #Creates blank labels at the end of chromosomes with tick
marks below axis

axis(1, at =Ref[1:31,1], labels =
c("","","","","","","","","","","","","","","","","","","","","","","","","","","","","","",""), tck
```

```
= -0.035, cex.axis = 0.8) #Creates blank labels at the end of chromosomes with tick
marks above axis

axis(1, at = RefC[1:npnts-1,1], labels =
c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,"X"),
tck = 0, cex.axis = 0.7) #Creates labels on the center of the chromosomes without tick
marks

legend("topright", legend = c("BayesB (0.997)", "BayesC (0.997)"), col =
c("gray70","black"), pch = c(16,16))

mtext("A)",side=3,line=1,adj=0)


###Warner-Bratzler Shear Force###
###BayesC, pi=0###
plot(xrange,yrange,yaxt="n",xaxt="n",xlab="Chromosome",ylab="PPAw2",cex.lab=0.7,
type="n")

for(i in 1:(npnts-1)){
	Chromo<-subset(WBSF, chr==i)
	nWin<-dim(Chromo)[1]
	for(j in 1:nWin){
		points(Chromo[j,1],Chromo[j,8],pch=16,col="black")
	}
}
axis(2,las=1,cex.axis=0.7)

axis(1,at=Ref[1:31,1],labels=c("","","","","","","","","","","","","","","","","","","","","","
","","","","","","","","",""),tck=0.035,cex.axis=0.8)
axis(1,at=Ref[1:31,1],labels=c("","","","","","","","","","","","","","","","","","","","","","
","","","","","","","","",""),tck=-0.035,cex.axis=0.8)
axis(1,at=RefC[1:npnts-
1,1],labels=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,2
8,29,"X"),tck=0,cex.axis=0.7)
#legend("topright", legend=c("BayesC (0.0)"), col="black",pch=16)
mtext("B)",side=3,line=1,adj=0)
abline(h = 0.75,lty=2)


###Overall Temperament At Weaning###
###BayesC, pi=0###
plot(xrange,yrange,yaxt="n",xaxt="n",xlab="Chromosome",ylab="PPAw2",cex.lab=0.7,
type="n")
```

```
for(i in 1:(npnts-1)){
	Chromo<-subset(OvWnDp, chr==i)
	nWin<-dim(Chromo)[1]
	for(j in 1:nWin){
		points(Chromo[j,1],Chromo[j,8],pch=16,col="black")
	}
}
axis(2,las=1,cex.axis=0.7)
axis(1,at=Ref[1:31,1],labels=c("","","","","","","","","","","","","","","","","","","","","","","
","","","","","","","","",""),tck=0.035,cex.axis=0.8)
axis(1,at=Ref[1:31,1],labels=c("","","","","","","","","","","","","","","","","","","","","","","
","","","","","","","","",""),tck=-0.035,cex.axis=0.8)
axis(1,at=RefC[1:npnts-
1,1],labels=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,2
8,29,"X"),tck=0,cex.axis=0.7)
#legend("topright", legend=c("BayesC (0.0)"), col=c("black"),pch=c(16))
mtext("A)",side=3,line=1,adj=0)
abline(h = 0.75,lty=2)
```

Objective 1 R Script: "Mapping Markers"

"GetGeneList" function. This portion of the R script takes user input and downloads the specified species gene information based on the specified map build. It further organizes and reduces the gene information based on the user's input. Examples are given at the bottom of the script. The user must run the function's code in R before calling it with the example.

```
GetGeneList =
function(Species,build,featuretype=c("GENE","PSEUDO"),savefiles=FALSE,destfile){
	if(missing(Species)){
		stop("ERROR: No species specified")
	}
	if(missing(build)){
		stop("ERROR: No map build specified")
	}
	if(missing(destfile)){
stop("ERROR: No path was specified for saving temporary and permanent files.")
	}
	cat("Please be patient, this could take a few minutes.","\n")
	Species = sub(" ","_",Species)
	dest = paste(destfile,"seq_gene.md.gz",sep="")
```

```
URL =
paste("ftp://ftp.ncbi.nih.gov/genomes/MapView/",Species,"/sequence/BUILD.",build,"/i
nitial_release/seq_gene.md.gz",sep="")
        download.file(URL,dest,cacheOK=TRUE)
        cat("Reading in file...","\n")
        NCBIList = read.table(gzfile(dest),header=FALSE,fill=TRUE)
        if(ncol(NCBIList) > 15){
                NCBIList<-NCBIList[,1:15]
        }
colnames(NCBIList) =
c("tax_id","chromosome","chr_start","chr_stop","chr_orient","contig","ctg_start","ctg_st
op","ctg_orient","feature_name","feature_id","feature_type","group_label","transcript","
evidence_code")
        ListF = matrix(0,1,15,dimnames=list(1,colnames(NCBIList)))
        for(j in featuretype){
                        List = subset(NCBIList,feature_type==j)
                        ListF = rbind(ListF,List)
        }
        ListF = ListF[2:nrow(ListF),]
        if(savefiles == TRUE){
write.table(NCBIList,paste(destfile,"seq_gene.txt",sep=""),quote=FALSE,sep="
",row.names=FALSE)
                remove(NCBIList)
        }
        if(savefiles == FALSE){
                unlink(dest)
                remove(NCBIList)
        }
        Assembly<-matrix(unique(ListF[,'group_label']),ncol=1)
        Features<-matrix(unique(ListF[,'feature_type']),ncol=1)
        Genes<-matrix(unique(ListF[,'feature_name']),ncol=1)
cat("Duplicate gene information may be present due to multiple assemblies and feature
types.","\n", "The following assembly builds are present in this gene list:","\n")
        print(Assembly)
        cat("The following feature types are present in this gene list:","\n")
        print(Features)
y = readline("Please choose which ASSEMBLY that you want to prioritize gene
information from \n (e.g. 1 for the first assembly listed, 2 for the second, etc.). \n Other
duplicate gene information (if any) will be removed from the list. \n ")
        y = as.numeric(y)
        if(abs(y) > nrow(Assembly)){
stop("ERROR: You specified a number outside the range possible for the assemblies.")
                }
                if(nrow(Features) > 1){
```

```
x = readline("Do you want to keep multiple feature type information? y = yes, n = no \n
")
                }else{ x = "y" }
                        if(x == "n"){
z = readline("Please choose which FEATURE TYPE that you want to prioritize gene
information from \n (e.g. 1 for the first feature listed, 2 for the second, etc.). \n Other
duplicated gene information will be removed from the list. \n")
                        z = as.numeric(z)
                        if(abs(z) > nrow(Features)){
stop("ERROR: You specified a number outside the range possible for the feature
types.")
                        }
                }
        ListF <- subset(ListF,group_label==Assembly[y,1])
        if(x == "n"){
                GeneList = subset(ListF,feature_type==Features[z,1])
        }
        if(x == "y"){
                GeneList = ListF
        }
        if(x != "y"){
                if(x != "n"){
stop("ERROR: You did not answer if you wanted duplicate feature type information
removed.","\n", "Please start over and enter y for yes or n for no when prompted.")
                }
        }
        if(savefiles==TRUE){
write.table(GeneList,paste(destfile,"GeneList.txt",sep=""),quote=FALSE,sep="
",row.names=FALSE)
        }
cat("Finished processing features and assemblies. The list will now be returned to the
user.","\n")
        return(GeneList)
}


##Examples
GeneList = GetGeneList("Bos taurus", build=6.1, savefiles=TRUE,
destfile="C:/Temp/")

GeneList = GetGeneList("Homo sapiens", 37.3, featuretype="RNA", savefiles=TRUE,
destfile="C:/Temp/")
```

```
GeneList = GetGeneList("Acyrthosiphon pisum", 2.1, savefiles=TRUE,
destfile="C:/Temp/")
```

"MapMarkers" function. This portion of the R script takes the output from the "GetGeneList" function with the marker map file provided by the user and maps the markers to the nearest gene, saves the distance and provides a category for whether the marker falls within the gene, nearby, or is far away. Although the function is coded to take input from the "GetGeneList" function, the user can simply provide the gene list that is to be used instead of using the previously described function. Below is the R code followed by an example. Please note that the code for the function must be run before the function can be called.

```
MapMarkers = function(features,markers,nAut,other=c("X"),destfile,savefiles=TRUE){
        if(missing(features)){
stop("ERROR: Did not specify list of features to use for mapping.")
        }
        if(missing(markers)){
                stop("ERROR: Did not specify list of markers to be mapped.")
        }
        if(missing(nAut)){
stop("ERROR: Did not specify the number of autosomes present in the marker file.")
        }
        if(savefiles = = TRUE){
                if(missing(destfile)){
stop("ERROR: No path was specified for the folder to save the output file.")
                }else{
                        dest = paste(destfile,"MappedMarkers.txt",sep="")
                }
        }
        if(other = = FALSE){
                chr = matrix(1:nAut, ncol=1)
        } else{
                Aut = matrix(1:nAut, ncol=1)
                nchr = nAut
                other = matrix(other, ncol=1)
                Oth = matrix(0,dim(other)[1],1)
                for(i in 1:nrow(other)){
                        Oth[i,1] = nchr+1
                        nchr = nchr+1
                }
                chr = rbind(Aut,Oth)
        }
```

```
Colnames = matrix(c(colnames(markers), colnames(features), "Distance", "Inside?"),
nrow=1)
                nCol = ncol(Colnames)
MarkMap = matrix(0, 1, nCol, byrow=TRUE, dimnames = list(c(1), Colnames))
                GLnCol = ncol(features)
                for(i in 1:nchr){
                        if(i > nAut){
                                j = i-nAut
                                k = other[j,1]
                        } else { k = i }
                        Chr_Features = subset(features, chromosome = = k)
                        Chr_Markers = subset(markers, chromosome = = i)
                        nmarkers = nrow(Chr_Markers)
                        nfeatures = nrow(Chr_Features)
                        rownames(Chr_Markers) = 1:nmarkers
                        rownames(Chr_Features) = 1:nfeatures
                        for(locus in 1:nmarkers){
                                MarkerInfo = subset(Chr_Markers[locus,])
                                MapPos = MarkerInfo[1,'position']
FeatureInfo = matrix(0, 1, GLnCol, byrow=TRUE, dimnames =
list(1,colnames(features)))
Inside = matrix(0, 1, 2, byrow=TRUE, dimnames = list(1, c("Distance", "Inside?")))
                                MapIt = cbind(MarkerInfo, FeatureInfo, Inside)
                                MaxDis = 1000000
                                for(feature in 1:nfeatures){
                                        Start = Chr_Features[feature,'chr_start']
                                        Stop = Chr_Features[feature,'chr_stop']
                                        DisStart = MapPos-Start
                                        DisStop = MapPos-Stop
                                        if(DisStart >= -2500){
                                                if(DisStop <= 0){
                                                        if(DisStart >= 0){
FeatureInfo = Chr_Features[feature,]
                                                                Inside[1,] = c(0,"Yes")
MapIt = cbind(MarkerInfo, FeatureInfo, Inside)
                                                        }
                                                        if(DisStart < 0){
FeatureInfo = Chr_Features[feature,]
Inside[1,] = c(abs(DisStart), "Close,_Before_Start_Position")
MapIt = cbind(MarkerInfo, FeatureInfo,Inside)
                                                        }
                                                }
                                        }
                                        if(DisStop <= 2500){
```

```
                                        if(DisStart >= 0){
                                                if(DisStop <= 0){
FeatureInfo = Chr_Features[feature,]
                                                        Inside[1,] = c(0,"Yes")
MapIt = cbind(MarkerInfo, FeatureInfo, Inside)
                                                }
                                                if(DisStop > 0){
FeatureInfo = Chr_Features[feature,]
Inside[1,] = c(abs(DisStop), "Close,_After_Stop_Position")
MapIt = cbind(MarkerInfo, FeatureInfo, Inside)
                                                }
                                        }
                                }
                        }
                        if(MapIt[,'Inside?'] == 0){
                                for(feature in 1:nfeatures){
                                        Start = Chr_Features[feature,'chr_start']
                                        Stop = Chr_Features[feature,'chr_stop']
                                        DisStart = MapPos-Start
                                        DisStop = MapPos-Stop
                                        if(DisStop <= 5000){
                                                if(DisStop > 2500){
FeatureInfo = Chr_Features[feature,]
Inside[1,] = c(abs(DisStop), "Far,_After_Stop_Position")
MapIt = cbind(MarkerInfo, FeatureInfo, Inside)
                                                }
                                        }
                                        if(DisStart >= -5000){
                                                if(DisStart < -2500){
FeatureInfo = Chr_Features[feature,]
Inside[1,] = c(abs(DisStart), "Far,_Before_Start_Position")
MapIt = cbind(MarkerInfo, FeatureInfo, Inside)
                                                }
                                        }
                                }
                        }
                        if(MapIt[,'Inside?'] == 0){
                                for(feature in 1:nfeatures){
                                        Start = Chr_Features[feature,'chr_start']
                                        Stop = Chr_Features[feature,'chr_stop']
                                        DisStart = MapPos-Start
                                        DisStop = MapPos-Stop
                                        if(DisStart >= -25000){
                                                if(DisStart < -5000){
```

```
FeatureInfo = Chr_Features[feature,]
Inside[1,] = c(abs(DisStart), "Very_Far,_Before_Start_Position")
MapIt = cbind(MarkerInfo, FeatureInfo, Inside)
                                                }
                                        }
                                        if(DisStop <= 25000){
                                                if(DisStop > 5000){
FeatureInfo = Chr_Features[feature,]
Inside[1,] = c(abs(DisStop), "Very_Far,_After_Stop_Position")
MapIt = cbind(MarkerInfo, FeatureInfo, Inside)
                                                }
                                        }
                                }
                        }
                        if(MapIt[,'Inside?'] == 0){
                                for(feature in 1:nfeatures){
                                        Start = Chr_Features[feature,'chr_start']
                                        Stop = Chr_Features[feature,'chr_stop']
                                        DisStart = MapPos-Start
                                        DisStop = MapPos-Stop
                                        if(DisStart > (-1*MaxDis)){
                                                if(DisStart < -25000){
FeatureInfo = Chr_Features[feature,]
Inside[1,] = c(abs(DisStart), "Nearest_gene_is_>_25,000_bp_after_marker")
MapIt = cbind(MarkerInfo, FeatureInfo, Inside)
                                                        MaxDis = abs(DisStart)
                                                }
                                        }
                                        if(DisStop < MaxDis){
                                                if(DisStop > 25000){
FeatureInfo = Chr_Features[feature,]
Inside[1,] = c(abs(DisStop), "Nearest_gene_is_>_25,000_bp_before_marker")
MapIt = cbind(MarkerInfo, FeatureInfo, Inside)
                                                        MaxDis = DisStop
                                                }
                                        }
                                }
                        }
                        if(MapIt[,'Inside?'] == 0){
                                MaxDis = 30000000
                                for(feature in 1:nfeatures){
                                        Start = Chr_Features[feature,'chr_start']
                                        Stop = Chr_Features[feature,'chr_stop']
                                        DisStart = MapPos-Start
```

```
                                                DisStop = MapPos-Stop
                                                if(DisStart > (-1*MaxDis)){
                                                        if(DisStart < -1000000){
FeatureInfo = Chr_Features[feature,]
Inside[1,] = c(abs(DisStart), "Nearest_gene_is_>_1,000,000_bp_after_marker")
MapIt = cbind(MarkerInfo, FeatureInfo, Inside)
                                                                MaxDis = abs(DisStart)
                                                        }
                                                }
                                                if(DisStop < MaxDis){
                                                        if(DisStop > 1000000){
FeatureInfo = Chr_Features[feature,]
Inside[1,] = c(abs(DisStop), "Nearest_gene_is_>_1,000,000_bp_before_marker")
MapIt = cbind(MarkerInfo, FeatureInfo, Inside)
                                                                MaxDis = DisStop
                                                        }
                                                }
                                        }
                                }
                        MarkMap = rbind(MarkMap,MapIt)
                        }
                }
                MarkMapF = MarkMap[2:nrow(MarkMap),]
                if(savefiles == TRUE){
                write.table(MarkMapF,dest,quote=FALSE,sep=" ",row.names=FALSE)
                }
                return(MarkMapF)
}

###Example###
GeneList = read.table("PlacePathHere/GeneList.txt",header=TRUE)

MarkerList = read.table("PlacePathHere /MarkerFile.txt", header=TRUE)

Test = MapMarkers(GeneList,MarkerList,29,other="X",destfile="C:/Temp/")
```

Objective 3 R Script: Handling Missing Breed-of-Origin Genotypes

```
Data = read.table("PlacePathHere.txt", header=FALSE)

nanimals = nrow(Data)
nmarkers = ncol(Data)-1
colnames(Data) = c("ID",1:nmarkers)

Families = array(c(70,71,72,73,74,75,76,77,80,81,82,83,84))
nuFamilies = dim(Families)

FamilyID = matrix(trunc(Data[,1]/100),nanimals,1)
Data2 = cbind(Data[,1],cbind(FamilyID,Data[,2:(nmarkers+1)]))
colnames(Data2) = c("ID","Family",1:nmarkers)
ncolumns = ncol(Data2)
remove(Data)

###Calculating family averages per genotype###

FamAvgGeno = matrix(0,nuFamilies,nmarkers)
for(i in 1:nuFamilies){
        DataTemp = Data2[which(Data2$Family == Families[i]),]
        FamAvgGeno[i,] = round(colMeans(DataTemp[,3:ncolumns],na.rm=TRUE),0)
}
FamAvgGeno = cbind(Families,FamAvgGeno)
colnames(FamAvgGeno) = c("Family",1:nmarkers)
write.table(FamAvgGeno,"PlacePathHere.txt",quote=FALSE,sep="
",row.names=FALSE)
remove(FamAvgGeno)


###Calculating average genotype across a family###

Data = read.table("PlacePathHere.txt",header=TRUE)
Data2 = replace(Data[,2:ncol(Data)],c(3,4),NA)

markermeans = rowMeans(Data2,na.rm=TRUE)
MeansByFamily = cbind(Data[,1],markermeans)
markermeans
mean(markermeans)


###Filling in unknown regions as same or heterozygotes for family average
genotypes###
```

```
AvgGenos = read.table("PlacePathHere.txt",header=TRUE)
for(i in 1:nuFamilies){
	temp = AvgGenos[i,2:(nmarkers+1)]
	colnames(temp) = c(1:nmarkers)
	loci = array(as.numeric(colnames(temp[,temp%in%c(3,4,'NA')])))
	nUnk = dim(loci)
	start = loci[1]
	k=1
	for(j in 2:nUnk){
		arrayend = loci[j+1]
		if(is.na(arrayend)==TRUE){
			end = loci[j]
if(((temp[,(start-1)] == 0)|(temp[,(start-1)]==1)|(temp[,(start-1)]==2))){
				temp[,start:end] = temp[,(start-1)]
				Same = "End,Same"
			} else {
				temp[,start:end] = 1
				Same = "End,Replacement=1"
			}
			sink("PlacePathHere.txt",append=TRUE)
cat("Family:",Families[i],"Start_is:",start,"End_is:",end,"The_loci_before_and_after_thi
s_region_were",Same,"\n")
			sink()
		} else {
			if((start+k) == loci[j]){
				k = k+1
				next
			} else {
				end = loci[j-1]
				if(start != 1){
if(((temp[,(start-1)] == 0)|(temp[,(start-1)]==1)|(temp[,(start-1)]==2)) &
((temp[,(end+1)] == 0)|(temp[,(end+1)]==1)|(temp[,(end+1)]==2))){
					if(temp[,(start-1)] == temp[,(end+1)]){
						temp[,start:end] = temp[,(start-1)]
						Same = "Same"
					} else {
						temp[,start:end] = 1
						Same = "Different,Replacement=1"
					}
				} else {
					temp[,start:end] = 1
					Same = "??Replacement=1"
				}
```

```
sink("PlacePathHere.txt",append=TRUE)
cat("Family:",Families[i],"Start_is:",start,"End_is:",end,"The_loci_before_and_after_thi
s_region_were",Same,"\n")
                                        sink()
                                        start = loci[j]
                                        k = 1
                                        } else {
if(((temp[,(end+1)] == 0)|(temp[,(end+1)]==1)|(temp[,(end+1)]==2))){
                                                temp[,start:end] = temp[,(end+1)]
                                                Same = "End,Same"
                                        } else {
                                                temp[,start:end] = 1
                                                Same = "Different,Replacement=1"
}
                                        sink("PlacePathHere.txt",append=TRUE)
cat("Family:",Families[i],"Start_is:",start,"End_is:",end,"The_loci_before_and_after_thi
s_region_were",Same,"\n")
                                        sink()
                                        start = loci[j]
                                        k = 1
                                }
                        }
                }
        }
        AvgGenos[i,2:(nmarkers+1)] = temp
}
colnames(AvgGenos) = c("Family",1:nmarkers)
write.table(AvgGenos,"PlacePathHere.txt",quote=FALSE,sep=" ",row.names=FALSE)


###Filling in Individual Animal's missing regions with family averages or same###

Genos = matrix(0,1,ncolumns)
colnames(Genos) = colnames(Data2)

for(i in 1:nanimals){
        temp = Data2[i,3:ncolumns]
        loci = array(as.numeric(colnames(temp[,temp%in%c(3,4,'NA')])))
        nUnk = dim(loci)
        start = loci[1]
        k = 1
        for(j in 2:nUnk){
                arrayend = loci[j+1]
                if(is.na(arrayend)==TRUE){
```

```
                        end = loci[j]
if(((temp[,(start-1)] == 0)|(temp[,(start-1)]==1)|(temp[,(start-1)]==2))){
                                temp[,start:end] = temp[,(start-1)]
                                Same = "End,Same"
                        } else {
info = matrix(AvgGenos[which(FamAvgGeno[,1] == Data2[i,2]),],1)
                                temp[,start:end] = info[,(start+1):end]
                                Same = "End,Different"
                        }
                        sink("PlacePathHere.txt",append=TRUE)
cat("Animal:",Data2[i,1],"Start_is:",start,"End_is:",end,"The_loci_before_and_after_this
_region_were",Same,"\n")
                        sink()
                } else {
                        if((start+k) == loci[j]){
                                k = k+1
                                next
                        } else {
                                end = loci[j-1]
                                if(start != 1){
if(((temp[,(start-1)] == 0)|(temp[,(start-1)]==1)|(temp[,(start-1)]==2)) &
((temp[,(end+1)] == 0)|(temp[,(end+1)]==1)|(temp[,(end+1)]==2))){
                                                if(temp[,(start-1)] == temp[,(end+1)]){
                                                        temp[,start:end] = temp[,(start-1)]
                                                        Same = "Same"
                                                } else {
info = matrix(AvgGenos[which(AvgGenos[,1] == Data2[i,2]),],1)
temp[,start:end] = info[,(start+1):(end+1)]
                                                        Same = "Different"
                                                }
                                        }
                                        sink("PlacePathHere.txt",append=TRUE)

cat("Animal:",Data2[i,1],"Start_is:",start,"End_is:",end,"The_loci_before_and_after_this
_region_were",Same,"\n")
                                        sink()
                                        start = loci[j]
                                        k = 1
                                } else {
if(((temp[,(end+1)] == 0)|(temp[,(end+1)]==1)|(temp[,(end+1)]==2))){
                                                temp[,start:end] = temp[,(end+1)]
                                                Same = "End,Same"
                                        } else {
info = matrix(AvgGenos[which(AvgGenos[,1] == Data2[i,2]),],1)
```

```
                                        temp[,start:end] = info[,(start+1):(end+1)]
                                        Same = "End,Different"
                                    }
                                    sink("PlacePathHere.txt",append=TRUE)
cat("Animal:",Data2[i,1],"Start_is:",start,"End_is:",end,"The_loci_before_and_after_this
_region_were",Same,"\n")
                                    sink()
                                    start = loci[j]
                                    k = 1
                                }
                            }
                        }
                    }
                    Genos = rbind(Genos,cbind(Data2[i,1:2],temp))
}

Genos = Genos[2:nrow(Genos),]
write.table(Genos,"C:/Perl64/bin/Cycle1_BoO_Fixed.txt",quote=FALSE,sep="
",row.names=FALSE)
```